NUCLEAR MAGNETIC RESONANCE STUDIES OF INSULIN-LIKE GROWTH FACTOR-I ANALOGUES DISPLAYING ENHANCED BIOLOGICAL ACTIVITY

by

Leanne G. Laajoki

B.Sc. (Hons) University of Wollongong

A thesis submitted for the degree of Doctor of Philosophy

of

The Australian National University

January 2000

Declaration

The work described in this thesis is original and has not previously been submitted for a degree or diploma in any other University or College, and to the best of my knowledge, does not contain material previously published or presented by another person, except where due reference is made in the text.

.....

Leanne G Laajoki

Acknowledgments

I would like to thank my supervisor, Dr Max Keniry for his patience, willingness to listen and always giving a large amount of his time and thought to my project. I would also like to thank Associate Professor John Carver, my co-supervisor for his enthusiasm, support and many helpful discussions. Thank you also to the other two members of my supervisory panel, Dr Nick Dixon, for his interest and offering the use of his lab; and Dr Andrew Torda, for always being "Bright and Cheery". I also thank Chris Blake and Elisabeth Owen for their help.

I was the grateful recipient of an Australian Postgraduate Award from the Australian Research Council. I would also like to acknowledge the Cooperative Research Centre for Tissue Growth and Repair in Adelaide, from whom I also received a scholarship. The people in the CRC directly involved with my work; Geoffrey Francis, Dr John Wallace and Dr Steve Milner, have all been very enthusiastic and patient. Thank you also to Gropep for the use of facilities and the support of staff members during my stay.

I would also like to express my appreciation for the efforts and support provided by Tony Herlt with HPLC, Penny Lilley for finding a home for me in the Dixon lab, Professor Peter Jeffrey for performing sedimentation equilibrium experiments, Dr John McLeod and Carl Braybrook for running mass spectra and Dr Sue Brown for help with BIAcore analyses. Thank you also to Dr Paul Gooley, Dr Gary Shooter, Dr Andrew Kralicek, Dr Grant Booker and Dr Terry Mulhern for helpful discussions.

To those I shared a room with, Dan Ayers, Adrian Cootes, Anthony Russell, Jim Proctor and Zsuzsanna Dosztany many thanks for the valuable friendships, the endless supply of jokes, and hastening a caffeine addiction.

Petri, thank you for your love, patience and tolerance. I am so fortunate that you can listen, hear and understand me. Thanks to my parents for being supportive of my aspirations and providing the opportunities for me to pursue my goals. Many thanks also to Jaakko and Anneli for their interest and excitement in everything I do. To all my friends who allow themselves to understand, thank you for your interest.

I dedicate this thesis to our Ben koira whose loyalty, motivation and love of the simple things in life inspire me. Also to Molly, may you always value your imagination and dreams.

Abstract

This research involves the NMR studies of two analogues of Insulin-Like Growth Factor-I (IGF-I) proteins, ^{15}N-labelled Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I. Both of these analogues have a thirteen amino acid extension appended to the N-terminus. In addition, Long-[Arg^3]-IGF-I has been modified by a Glu^{3}Arg mutation while Long-[Leu^{60}]-IGF-I has been modified by a Tyr^{60}Leu mutation. For both analogues, the resonance assignments, secondary structure, amide proton exchange rates, $^3J_{NH\alpha}$ coupling constants, and backbone dynamics have been determined using high resolution NMR spectroscopy. For Long-[Arg^3]-IGF-I, the three dimensional solution structure has been determined using restrained molecular dynamics. Chemical shift mapping studies of Long-[Leu^{60}]-IGF-I have been performed by forming a complex between Long-[Leu^{60}]-IGF-I and insulin-like growth factor binding protein 2 (IGFBP-2).

In both analogues, the secondary structural elements of IGF-I, three α-helices forming a hydrophobic core, have been retained. In Long-[Arg^3]-IGF-I, some variations in the structure are noted in the N-terminus and a slight reorientation of helices 2 and 3 relative to helix 1. These results are interpreted in terms of lower binding affinity for the IGFBPs. The N-terminal extension of Long-[Arg^3]-IGF-I has very few NOE contacts either to the IGF-I domain or within the extension. However, many amide protons of the N-terminal extension are slowly exchanging, leading to the postulate of a molten helical structure for this extension. The backbone dynamics analysis indicate that there is a considerable degree of flexibility in both Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I, leading to an anomalously high estimate for the overall correlation time. The highest order parameters are observed in the helical regions and the lowest order parameters in the C-domain loop separating helices 1 and 2 and in the N-terminal extension. There are some subtle differences in the order parameters between the two analogues, particularly for the residues following α-helix 1, a region implicated in binding to the type 1 IGF receptor.

There are differences in the chemical shifts of the ^{15}N, H^N and H^α nuclei for the two analogues, in the proximity of residue 60 and the N-terminal extension as well as the C-domain and α-helix 1 which are well separated in primary sequence from residue 60 but close in three dimensional space. There are quite substantial changes in the solvent accessibility between the two proteins, with approximately half the number of slowly exchanging amide protons in Long-[Leu^{60}]-IGF-I. Whilst many of the amide protons in

the N-terminal extension are slowly exchanging, Leu^{60} is the only helical residue with an amide proton observed in the 2D ^{15}N-HSQC spectra after 24 hours.

Many Long-[Leu^{60}]-IGF-I residues were implicated by chemical shift mapping experiments at the binding interface with IGFBP-2. These residues were located primarily in the N-terminus of the IGF-I domain of the protein and the C-terminal end of α-helix 2. Some resonances from residues in the less well defined N- and C-termini were also perturbed upon binding of IGFBP-2, likely a result of indirect conformational changes.

Overall, these investigations contribute to the understanding of the structure and mobility of the IGF family and their interactions with IGF receptors and IGFBPs.

Contents

List of Figures

List of Tables

List of Abbreviations

ACTH	Adrenocorticotrophic Hormone
ALS	Acid Labile Subunit
ANGIS	Australian National Genomic Information Service
ANU	Australian National University
AQUA	Program for Analyzing the QUAlity of biomolecular structures determined by NMR spectroscopy
AUFS	Absorbance Units Full Scale
BMRB	BioMagResBank
C4	Hydrocarbon Chain, ie. butyl
CALIBA	Calibrating NOEs (macro)
CD	Circular Dichroism
CM5	Carboxymethyl sensor chip
CPMG	Carr-Purcell-Meiboom-Gill
CRC	Cooperative Research Centre
CSA	Chemical Shift Anisotropy
CSI	Chemical Shift Index
C-terminal	Carboxy Terminal
2D	Two Dimensional
3D	Three Dimensional
DD	Dipole-Dipole contribution to transverse relaxation
DG	Distance Geometry
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol

DYANA	Dynamics Algorithm for NMR Applications
EDTA	Ethylenediaminetetra-acetic acid
FFS	Fast Flow Sepharose
FID	Free Induction Decay
FSH	Follicle Stimulating Hormone
FTIR	Fourier Transform Infrared Spectrometry
GH	Growth Hormone
GLUT4	Glucose Transporter 4
HBS	HEPES Buffer Saline
HEPES	N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
HFBA	Heptaflourobutyric acid
HMQC	Heteronuclear Multiple Quantum Coherence
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Coherence
IGF-I	Insulin-Like Growth Factor 1
IGF-II	Insulin-Like Growth Factor 2
IGFBP	Insulin-Like Growth Factor Binding Protein
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
IPTG	Isopropyl β-D-thiogalactopyranoside
K_D	Dissociation Constant
MSA	Multiple Stimulating Activity
NMR	Nuclear Magnetic Resonance Spectroscopy
NOE	Nuclear Overhauser Effect

NOESY	Nuclear Overhauser Effect Spectroscopy
NSILA	Non-suppressible Insulin-Like Activity
N-terminal	Amino Terminal
ppm	parts per million
pGH	porcine Growth Hormone
PDB	Protein Data Bank
RCM	Radial Compression Module
rMD	restrained Molecular Dynamics
RMSD	Root Mean Square Deviation
RNA	Ribonucleic Acid
rpm	Revolutions per minute
SGE	Scientific Glass Engineering
TFA	Triflouroacetic acid
TOCSY	Total Correlation Spectroscopy
UV	Ultraviolet
VNMR	Varian NMR analysis software

CHAPTER 1

Introduction

Insulin-like growth factors (IGFs) are small proteins that participate in the control of cell growth and replication. *In vivo,* IGFs are bound to insulin-like growth factor binding proteins (IGFBPs) which modulate the half life of IGFs in blood and assist in the transport and delivery of IGFs to their cellular receptors (reviewed by Walton et al., 1990).

1.1 Historical Perspective

The IGFs were discovered as a result of the contributions from three separate lines of research. In 1957, Salmon and Daughaday observed that growth hormone (GH), in isolation, does not stimulate growth processes, but rather, induces the biosynthesis of factors that mediate the messages of GH. These factors were initially referred to as "sulphation factors" and later as "somatomedins".

Following the development of a radioimmunoassay for insulin, it became apparent that serum contained far more insulin-like activity than could be accounted for simply by its content of immunoreactive insulin. In addition, the insulin effects of serum were not suppressed by the addition of anti-insulin serum, contrary to the effects of insulin. Froesch et al. (1963) termed this insulin-like activity "non suppressible insulin-like activity" (NSILA).

In 1972, Pierson and Temin extracted factors from serum which stimulated cells to replicate when added to culture medium. These factors were named "multiplication-stimulating activity" (MSA). Later, it was found that cultured liver cells secrete MSA into the culture medium, demonstrating that MSA is produced in the liver and that MSA may lead to autocrine stimulation of the same cells from which it is secreted.

Rinderknecht and Humbel, in 1976, purified and sequenced NSILA from a Cohn fraction of human plasma (Rinderknecht and Humbel, 1978). The primary sequence was found to be 48% homologous with human proinsulin. The sequence of a second insulin-like molecule was found to be similar. These biologically active peptides were termed IGF-I and IGF-II. Subsequent sequencing of purified somatomedin C from human plasma showed its sequence to be identical to that of IGF-1 (Svoboda, et al., 1980).

The biosynthesis of IGF-II is under less stringent control by GH than the biosynthesis of IGF-I (Humbel, 1990). Since the name somatomedin had been introduced to imply GH

regulation, it was an inadequate term to describe both of these homologous peptides. Thus in 1987, it was recommended that the term IGF be employed to refer to this family of peptides (Sara and Hall, 1990). The IGFs are members of the insulin superfamily, which includes insulin, IGF-I, IGF-II, relaxin, bombyxin, and the molluscan insect peptides (Murray-Rust et al., 1992).

1.2 IGF-I and IGF-II Sequences and Homology

IGF-I and IGF-II were characterized by Rinderknecht and Humbel (1976 and 1978). IGF-I and IGF-II are single chain polypeptides of 70 (mol. wt. 7649) and 67 (mol. wt. 7471) amino acids respectively, each with three disulfide bonds. The two proteins are 65% homologous with each other, while IGF-I and IGF-II are 43% and 41% homologous with insulin respectively.

The polypeptides consist of four domains; B, C, A, and D in sequential order from the N-terminus. The A and B domains are structurally homologous with the insulin A and B chains. Table 1.1 shows the sequences of human IGF-I, IGF-II and insulin. In IGF-I, residues 3-29 correspond to the B chain of insulin while residues 42-62 correspond to the insulin A chain. In IGF-II, residues 6-32 correspond to the insulin B chain while residues 41-61 correspond to the insulin A chain. The C domain is analogous to the connecting peptide in proinsulin, though it is quite different in sequence. The IGFs are the only members of the insulin superfamily in which the C region is not removed proteolytically after translation. Unlike proinsulin, the IGFs also have a carboxyl terminal extension of eight amino acids from residues 63 to 70 in IGF-I and residues 60 to 67 in IGF-II, which is known as the D domain. There are three disulfide bonds linking residue 6 with 48, 18 with 61, and 47 with 52 in IGF-I, while the corresponding linkages in IGF-II are 9 with 47, 21 with 60, and 46 with 51.

Table 1.1: Sequences of Human Insulin, IGF-I and IGF-II with the B, C, A and D Domains shown.

B	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Insulin			P	V	N	Q	H	L	C	G	S	H	L	V	E	A	L	Y	L	V	C
IGF-I				G	P	E	T	L	C	G	A	E	L	V	D	A	L	Q	F	V	C
IGF-II	A	Y	R	P	S	E	T	L	C	G	G	E	L	V	D	T	L	Q	F	V	C
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

	20	21	22	23	24	25	26	27	28	29	30
Insulin	G	E	R	G	F	F	Y	T	P	K	A
IGF-I	G	D	R	G	F	Y	F	N	K	P	T
IGF-II	G	D	R	G	F	Y	F	S	R	P	A
	19	20	21	22	23	24	25	26	27	28	29

C	30	31	32	33	34	35	36	37	38	39	40	41
IGF-I	G	T	G	S	S	S	R	R	A	P	Q	T
IGF-II	S	-	-	R	V	S	R	R	S	R	-	-

A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Insulin	G	I	V	E	Q	C	C	T	S	I	C	S	L	Y	Q	L	E	N	Y	C	N
IGF-I	G	I	V	D	E	C	C	P	R	S	C	D	L	R	R	L	E	M	Y	C	A
IGF-II	G	I	V	E	E	C	C	F	R	S	C	D	L	A	L	L	E	T	Y	C	A
	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62

D	63	63	65	66	67	68	69	70
IGF-I	P	L	K	P	A	K	S	A
IGF-II	T	-	-	P	A	K	S	A

The sequence homology of IGFs is well conserved across species as is demonstrated by the following comparisons. Bovine, porcine and human IGF-I are identical, whereas rat and mouse IGF-I differ from human IGF-I by three and four amino acids respectively. Human and bovine IGF-II differ by three amino acids, while rat and mouse IGF-II differ from human IGF-II by four and six amino acids respectively. These species differences in IGF-II occur toward the end of the B domain and within the C domain (Sara and Hall, 1990). The arginine residues of IGF-I, which are involved in binding interactions with other proteins (Jansson et al., 1998; Zhang et al., 1994), are highly conserved in vertebrates. The only common variant is the arginine to glutamine substitution at position 50 in chicken, frog, and salmon IGF-I.

1.3 Solution Structures of IGFs

Refined structures of proteins from the IGF family include two reports of the IGF-I structure (Cooke et al., 1991; Sato et al., 1992,1993), two structures of IGF-II (Torres et al., 1995; Terasawa et al., 1994) and the structure of a mini-IGF-I, in which the C-domain is deleted (DeWolf et al., 1996). The biological implications of the deletion of the C-domain of mini-IGF-I are discussed in section 1.8. No crystals suitable for X-ray analysis have been obtained for either IGF-I or IGF-II. The tertiary structures of IGFs have been modeled on that of porcine insulin (Blundell et al., 1978, 1983). Structures determined by NMR spectroscopy in solution are consistent with the modeled structure.

1.3.1 IGF-I

The secondary structure and three-dimensional structure of IGF-I were determined by ^{1}H NMR spectroscopy by Cooke et al. (1991) followed by Sato et al. (1992, 1993). In IGF-I there are three helical segments corresponding to the regions Ala^{8}-Cys^{18} (helix 1), Gly^{42}-Cys^{48} (helix 2) and Leu^{54}-Cys^{61}(helix 3). A β-turn, encompassing residues Gly^{19}-Gly^{22}, immediately follows helix 1. Residues Phe^{23} through to Thr^{41} exist in an extended structure with the possibility of some fixed conformation in the region Thr^{29}-Ser^{35}. The D domain is considered to be unstructured.

The spatial arrangement of the three helices of insulin is conserved in the human IGF-I structure, despite the C domain being inserted between the A and B domains. The three helical rods form a rigid structural core which may be stabilized through hydrophobic interactions. The C-domain is flexible and unstructured and forms an exposed loop. Figure 1.1 shows the average solution structure of native IGF-I with the spatial orientation of the three helices (Cooke et al., 1991). The axes of helices 2 and 3 are parallel to each other and roughly perpendicular to helix 1 (Sato et al., 1993).

1.3.2 IGF-II

The structure of IGF-II (Terasawa et al., 1994; Torres et al., 1995) in the well defined regions is very similar to those of the corresponding regions of insulin and IGF-I. The main secondary structure elements of IGF-II are α-helices encompassing residues 11-21 in the B domain, and residues 42-48 and 53-59 in the A domain. The two helices in the A domain are anti-parallel to each other and perpendicular to helix 1. Residues 6-9 and 25-

30 are in extended conformations. A reverse turn occurs at residues 22-25 leading into the extended region. The amide protons of residues 25-28, which are in an extended conformation, are all slowly exchanging and have downfield H^{α} chemical shifts relative to random coil values suggesting a β-sheet conformation. Thus, in IGF-II it appears that there is a small region of anti-parallel sheet formed by residues 25-27 and 59-61 immediately after helix 3 (Torres et al., 1995).

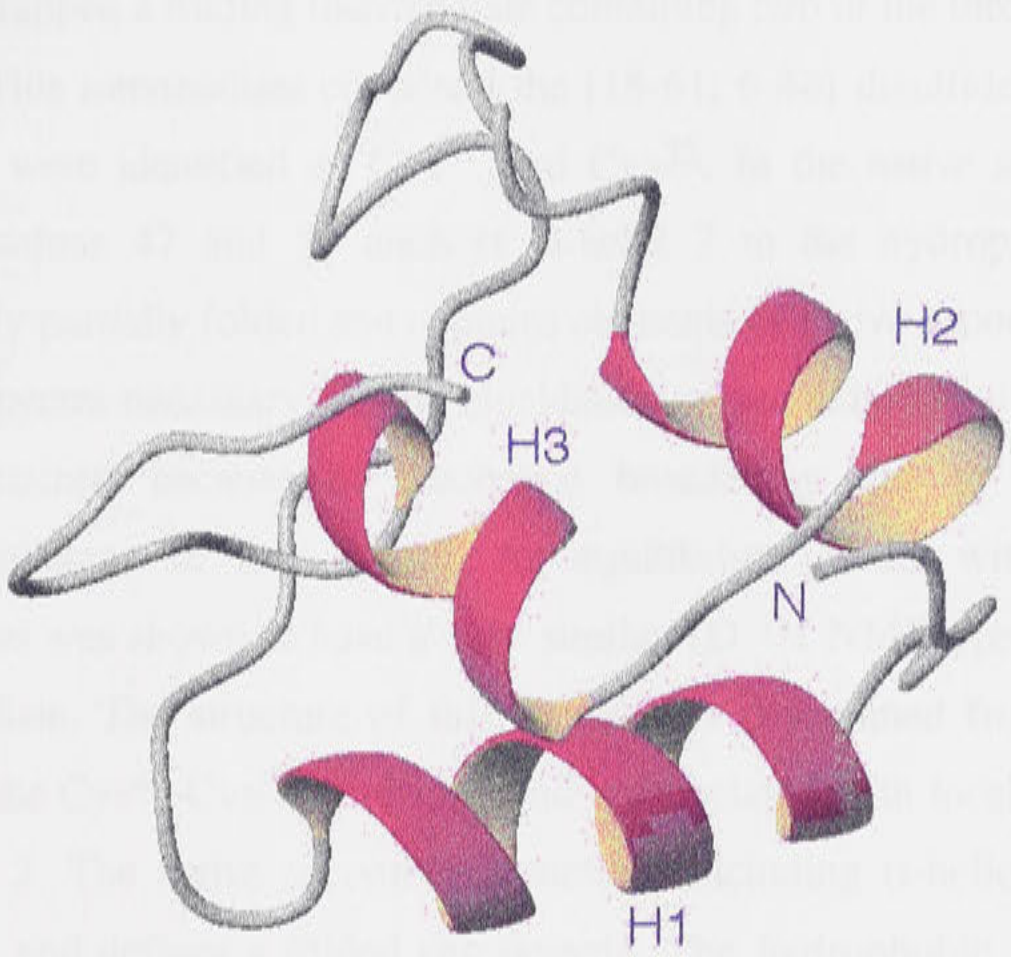

Figure 1.1: NMR solution structure of IGF-I (Cooke et al., 1991). The three α-helices are labeled H1, H2 and H3 respectively. The N- and C-termini are also labeled.

1.3.3 Mini-IGF-I

The secondary structure elements of IGF-I are retained in mini-IGF-I (De Wolf et al., 1996). The two structures, IGF-I and mini-IGF-I are similar globally, although the deletion of the C-domain results in differences in the spatial arrangement of the helices. The axes of helix 2 and helix 3 in mini-IGF-I are no longer parallel. The conformation of the segment encompassing residues Phe^{23}, Tyr^{24} and Phe^{25} differs from that of wild-type IGF-I. If the conformation of this segment were the same as that of wild-type IGF-I, the restrictions imposed by the deletion of the C-domain would result in exposure of residues of the hydrophobic core to the solvent, in particular Phe^{23} and Phe^{25}. However, a displacement of this segment, facilitated by the flexibility of the preceding β-turn (Gly^{19}-

Gly^{22}) allows the core to remain shielded and the important interactions of Phe^{23} and Phe^{25} with residues in the neighbouring helices to be maintained. It is postulated that this change in conformation of the segment Phe^{23}, Tyr^{24} and Phe^{25} may be the driving force for the movement of the helix 2-turn-helix 3 motif relative to helix 1.

1.3.4 Partially Reduced IGF-I and an Alanine Model

Hua et al. (1996) trapped a folding intermediate containing two of the three native disulfide bonds of IGF-I. This intermediate contained the [18-61; 6-48] disulfide bonds, while the reduced cysteines were identified as Cys^{47} and Cys^{52}. In the native state, the disulfide bond between residues 47 and 52 anchors α-helix 2 to the hydrophobic core. This intermediate is only partially folded and contains elements of native secondary and tertiary structure. NMR spectra necessary for the elucidation of secondary and tertiary structure could not be obtained because of resonance broadening due to the intermediate conformational exchange of this variant. An equilibrium model with Cys^{47}Ala and Cys^{52}Ala mutations was shown to have a very similar 1D ^{1}H NMR spectrum to the [18-61;6-48] intermediate. The structure of this model was determined from 2D ^{1}H NMR data. Removal of the Cys^{47}-Cys^{52} disulfide bond is associated with local unfolding of the adjoining α-helix 2. The native secondary structure (including α-helices 1 and 3) was otherwise retained and defines a folded sub domain. The hydrophobic core enclosed by helix 1, the receptor binding strand (Phe^{23} to Phe^{25}) and helix 3 was retained. Hua et al. (1996) proposed that the folding of the insulin motif is directed by a subset of native structural elements and that these elements form at an early step in the folding pathway of IGF-I. Formation of helix 2, despite its prominence in the native state, is likely to represent a late step in the folding process. Hydrophobic collapse of this segment may precede helix formation.

1.4 Gene Expression, Synthesis and Regulation

Single genes code for IGF-I and IGF-II (Rutanen and Pekonen, 1990). The IGF-I gene is located on the long arm of chromosome 12, containing five exons over a region of 85 kilobase pairs (kb) of genomic DNA (Walton et al., 1990). The human IGF-I gene is transcribed and processed into at least two types of messenger RNA (mRNA) (Rutanen and Pekonen, 1990). The IGF-II gene is located on the short arm of chromosome 11, 4.4

kb downstream from the insulin gene (Walton, et al., 1990). This gene contains eight exons, five non-coding and three protein encoding exons, and is transcribed into three mRNAs (Rutanen and Pekonen, 1990). IGF-II is synthesized as a 22 kDa prepropeptide and then post-translationally processed to a 20 kDa propeptide, prior to final processing to the mature IGF-II peptide (Walton et al., 1990).

IGF-I and IGF-II are synthesized in multiple organs and tissues. Most tissues express both IGFs at some stage of pre- or postnatal development (Rutanen and Pekonen, 1990). Han et al. (1987) demonstrated that expression of both IGF-I and IGF-II are localized to connective tissue or cells of mesenchymal origin. IGF-II dominates during the prenatal period (Rutanen and Pekonen, 1990). The liver represents the major site of expression and of synthesis during postnatal growth (Steele and Elsasser, 1989).

The major factor regulating IGF-I production is growth hormone (GH) (Mathews et al., 1986). GH stimulates the induction of IGF-I mRNA in various organs, tissues, primary cell cultures and in established cell lines. A rapid rise in serum IGF-I levels is seen after GH treatment of patients with GH deficiency (Rutanen and Pekonen, 1990). Oestrogen also stimulates IGF-I expression (Murphy et al., 1987; Ernst et al., 1989; and Johnson et al., 1989), while parathyroid hormone has been shown to stimulate IGF-I production in bone cells (Canalis et al., 1989). During fasting, protein malnutrition, and diabetes, IGF-I production and IGF-I mRNA expression in the liver is decreased. Serum IGF-I production is low in the human neonatal period, rises to peak values during puberty, and then slowly declines during aging (Hall and Sara, 1983).

The factors regulating IGF-II production are poorly characterized (Rutanen and Pekonen, 1990). IGF-II mRNA expression is not GH dependent. The highest levels of mRNA expression of IGF-II are detected in the brain and expression is barely detected in the liver (Murphy et al., 1987; and Hynes et al., 1987). Placental lactogen has a stimulatory effect on foetal fibroblast IGF-II production (Adams et al., 1983). IGF-II mRNA in humans is highest during foetal development and declines in the postnatal period (Rutanen and Pekonen, 1990). Follicle Stimulating Hormone (FSH) regulates IGF-II mRNA in human granulosa cells and ACTH in foetal adrenal cells (Voutilainen et al., 1987). In adult, high levels of IGF-II are found in bone, where it may be a mediator of remodelling, and in the central nervous system. High levels of IGF-II mRNA are also found in a number of tumours, including mesenchymal tumours and Wilms' tumours (Rechler et al., 1991).

IGF-II also has insulin-like metabolic effects, and overproduction of IGF-II in certain non-islet cell tumours results in refractory clinical hypoglycaemia (Daughaday and Rotwein, 1989). The insulin-dependent delivery of the IGF-II/Man-6-P receptor to the cell surface from the intracellular storage pool is mediated by the tissue-specific glucose transporter isoform vesicles (GLUT4) (Kandror and Pilch, 1996).

1.5 IGF Receptors

1.5.1 Introduction

The biological actions of IGFs are mediated by specific cell membrane receptors, mainly through the type 1 IGF receptor which is found on the surface of most cell types (Leroith et al., 1997; Ballard et al., 1994). The receptors that recognise the IGFs can be distinguished by their primary structure, relative affinities for their respective ligands, as well as by immunological methods (Rutanen and Pekonen, 1990). IGF-I and IGF-II bind to three types of membrane-bound receptors on target cells: the type 1 IGF receptor, the type 2 IGF receptor and the insulin receptor. Insulin cross reacts weakly with the type 1 IGF receptor but has no measurable affinity for the type 2 IGF receptor (Massagué & Czech, 1982). The type 1 IGF receptor has a high degree of sequence homology with the insulin receptor, while the type 2 IGF receptor is homologous with the cation-independent mannose-6-phosphate receptor, which functions in the trafficking of lysosomal enzymes but has no known IGF signalling function. The insulin receptor binds the IGFs with low affinity. Hybrid IGF/insulin receptors have been isolated that bind the IGFs and potentially transmit a cytoplasmic signal *in vitro* (Jones and Clemmons, 1995). Soos et al. (1990) demonstrated the existence of hybrid receptors in normal tissues of rat fibroblasts and showed that both insulin and IGF-I are bound with high affinity.

A chimeric type 1 IGF receptor in which the C-terminal domain was replaced with that of the insulin receptor was constructed by Tartare et al. (1994). This receptor possessed fully functional kinase activity and biological properties. The C-terminal domain of the insulin receptor is more tightly coupled to the stimulation of glycogen synthesis than is that of the type 1 IGF receptor. It was proposed that this domain plays a crucial role in the transmission of biological effects and could account, in part, for receptor specificity.

1.5.2 IGF Receptor and Insulin Receptor Structures

Specific growth factor receptors with high affinity for IGF-I and IGF-II were identified and characterized by Massagué and Czech (1982).

The type 1 receptor, like the insulin receptor, is a glycosylated heterotetramer composed of two α subunits and two β subunits linked by disulfide bonds. The α subunits consist of 706 amino acids (135 kDa in mass) and exist extracellularly while the β subunits are 90 kDa in mass (626 amino acids), and contain a transmembrane domain and the intracellular tyrosine autophosphorylation site. The α subunits contain the cysteine-rich, extracellular ligand binding site (Humbel, 1990). The subunits of the insulin receptor are very similar to the type 1 IGF receptor in size and have regions with a high degree of sequence homology (50-60% overall sequence identity and 84% in the tyrosine kinase domains) (Walton et al., 1990). The type 1 IGF receptor is synthesized on the ribosome as a single polypeptide chain of 1367 amino acid residues which is post-translationally modified by the removal of a 30 amino acid signal peptide and cleavage of the proreceptor into α and β subunits linked by disulfide bonds, which is referred to as an αβ-half receptor. The αβ-half receptor is subsequently linked to another αβ-half receptor to form the mature $\alpha_2\beta_2$-holoreceptor (Jones and Clemmons, 1995).

The type 2 IGF receptor is structurally unrelated to the type 1 IGF and insulin receptors, although, as mentioned above, it is homologous with the mannose-6-phosphate receptor which is believed to be involved in the transport of lysosomal enzymes (Walton et al., 1990). It is a single chain, 250 kDa polypeptide chain and has no intrinsic kinase activity (Rutanen and Pekonen, 1990). There is a large extracellular domain (approximately 93% of the protein) which contains 15 repeat sequences of cysteine residues and hydrophobic regions (Humbel, 1990), and a small cytoplasmic domain. Although the type 2 IGF receptor and the mannose-6-phosphate receptor are homologous, their functions are unrelated. The mannose-6-phosphate receptor is located on both the cell surface and the Golgi apparatus where it binds lysosomal proteins and mediates their transport to the lysosomes. The type 2 IGF receptor and the mannose-6-phosphate receptor also have distinct binding sites on the receptor for their different ligands. However, the two ligands appear to act synergistically since mannose-6-phosphate increases the affinity of the receptor for IGF-II (Jones and Clemmons, 1995). Murayama et al. (1990) also showed that the mannose-6-phosphate receptor has two distinct signalling functions that positively

or negatively regulate the activity of G_{i-2} (subunit of a guanidine nucleotide protein) in response to the binding of IGF-II or mannose-6-phosphate.

1.5.3 Ligand Binding

Type 1 and type 2 IGF receptors and insulin receptors are present on most cells in the body (Humbel, 1990). This arrangement has complicated the assignment of a particular response to a particular receptor. The nomenclature of the IGF receptors is based on their specificity for binding to each peptide. In most tissues, the type 1 IGF receptor binds IGF-I most strongly, followed by, with decreasing affinity, IGF-II and then insulin. The insulin receptor binds insulin more strongly than IGF-I and IGF-II with the latter two having similar binding affinities. The binding of IGF-I and insulin to the type 2 IGF-I receptor is very low and the association constant is biologically insignificant (Kiess et al., 1994). The IGFs themselves appear to undergo a conformational change on binding to their receptors (Gill et al., 1996; Jansson et al., 1998). Upon binding of IGFs to the type 1 IGF receptor, the intracellular part of the receptor undergoes autophosphorylation, which is a prerequisite for all known IGF-mediated biological actions. The growth promoting effects of insulin and the metabolic activity of IGF-I are thought to arise from the cross binding to each other's receptors.

Differences in affinities of IGF-I and IGF-II for the various receptors arise predominantly from differences in key sidechains rather than from structural differences between the two proteins. Cascieri et al. (1988) showed that the structural determinants of IGF-I required to maintain binding to the type 1 IGF receptor and to the type 2 IGF receptor and serum binding proteins are different. Cascieri et al. (1988) identified Tyr^{24} of IGF-I as an important residue for binding of the protein to the type 1 IGF receptor. Mutation of Tyr^{24} to the non-aromatic residues, Leu or Ser, resulted in a large decrease in type 1 IGF receptor affinity. Combination of this mutation with deletion of the D domain led to a further loss in receptor affinity. The mutation of IGF-I at Tyr^{24} also resulted in a loss of affinity for the insulin receptor, though it was not as marked as the loss in type 1 IGF receptor affinity. However, the alteration of IGF-I from the wild type Phe^{23}, Tyr^{24}, Phe^{25} to give the mutant [Phe^{23}, Phe^{24}, Tyr^{25}]IGF-I does not result in major changes in affinity for either the type 1 or type 2 IGF receptors, indicating that the tyrosine hydroxyl group at residue 24 does not have a direct effect on binding affinity. While the loss of aromaticity at residue 24 of IGF-I results in a loss of affinity for the type 1 IGF and insulin receptors,

there is little difference in affinity for the type 2 IGF receptor (Cascieri et al., 1988). Mutation of the aromatic residue at position 23 results in a loss in affinity for all receptors (Cascieri et al., 1988). Maly and Lüthi (1988) performed iodination experiments to elucidate the IGF-I binding site, involving three tyrosine residues, Tyr^{24}, Tyr^{31} and Tyr^{60}, and proposed that the minimal receptor binding region of IGF-I would include amino acid residues Tyr^{24} to Tyr^{31} and, possibly, the C-terminal part of the A-domain with Tyr^{60}. Bayne et al. (1988) mutated three tyrosine residues of IGF-I, Tyr^{24}, Tyr^{31} and Tyr^{60} to either an alanine or a leucine to generate either single, double or triple point mutations. These molecules showed reduced binding to the type 1 IGF receptor compared with the parent molecule.

The C domain of IGF-I has been shown to be important for the binding interaction with IGF receptors. Mini-IGF-I (deletion of the C domain residues 28-41) was found to have much lower affinity for the type 1 receptor (De Wolf et al., 1996). Gill et al. (1996) reported that mini-IGF-I has 100 times lower affinity than IGF-I for the type 1 IGF receptor. Bayne et al. (1988) prepared a mutant of IGF-I in which the C domain (residues 28-37) was replaced with a four glycine bridge. This mutant had a 30-fold loss of affinity for the type 1 IGF receptor, however deletion of the D domain (residues 63-70) had little effect on binding to the type 1 IGF receptor. Neither the C- or D-domains of IGF-I were required for high affinity binding to the type 2 IGF receptor. However, Zhang et al. (1994) have subsequently shown the positively charged residues of Arg^{36} and Arg^{37} in the C domain and Lys^{65} and Lys^{68} in the D domain contribute to receptor-binding specificity and high-affinity binding to the type 1 IGF receptor.

Blundell et al. (1978) have proposed that a number of invariant residues shared between IGF-I and insulin participate in IGF-I binding to the type 1 IGF receptor and its cross reactivity with the insulin receptor. Structural determinants in the C- and D-domains are involved in maintaining receptor binding specificity between the two receptors (Bayne et al., 1988; Zhang et al., 1994). The distinct physiological roles of IGF-I and insulin have been enhanced during evolution by variation in receptor signalling, receptor tissue distribution and ligand binding specificity. It is still unclear how the receptors differentiate between the physiological roles of IGF-I and insulin. A number of A- and B-domain positions in IGF-I that are important for maintaining high-affinity receptor binding and specificity have been identified (for example, Ala^{8}, Met^{59}, Ala^{62}) (Shooter et al., 1996). The receptor specificity of IGF-I evolved with a number of sequence changes within the

B- and A-domains that collectively contribute to the observed receptor binding properties of IGF-I (Shooter et al., 1996).

Cascieri et al. (1989) prepared IGF-I mutants with specific residues in the A-domain replaced with the corresponding residues in the A chain of insulin in order to investigate determinants for type 2 IGF receptor binding. Residues Phe^{49}, Arg^{50} and Ser^{51} were found to be important for maintaining binding to the type 2 IGF receptor, while binding to the type 1 IGF receptor was unaffected. Bayne et al. (1987) prepared an IGF-I mutant in which the 16 amino-terminal residues were replaced with the 17 amino-terminal residues of the insulin B chain, and found that this change causes a dramatic loss of binding affinity for the type 2 IGF receptor. While the binding determinants in IGF-I, important for type 2 IGF receptor binding, are different from those required for maintaining high affinity for the type 1 IGF receptor, this region in the A-domain overlaps with the determinants in IGF-I for high affinity binding to insulin-like growth factor binding proteins (IGFBPs) (discussed in section 1.6).

Sakano et al. (1991) identified the amino acids Phe^{48}, Arg^{49}, Ser^{50}, Ala^{54} and Leu^{55} as the residues in IGF-II which are critical for binding of IGF-II to the type 2 IGF/mannose-6-phosphate receptor. These residues form a patch when mapped on a three-dimensional structure of IGF-II (Terasawa et al., 1994). In contrast, the residues in IGF-II which are responsible for the binding of IGF-II to the type 1 IGF receptor or the insulin receptor are Phe^{26}, Tyr^{27} and Val^{43} (Sakano et al., 1991). Roth et al. (1991) also found Tyr^{27} to be an important determinant for type 1 IGF receptor binding, and proposed that the D-domain of IGF-I forms part of the binding region for the type 1 IGF receptor since it is close in space to Tyr^{27}.

1.5.4 Biological Responses Mediated by IGFs

In general, the insulin receptor has been implicated in rapid anabolic responses, while the mitogenic functions of the IGFs are almost always mediated by the type 1 IGF receptor (Humbel, 1990). However, insulin stimulates a proliferative response in certain cells through its own receptor, and the type 1 IGF receptor can, in some cells, mediate rapid anabolic responses. Most of the responses stimulated by IGF-II are thought to be mediated by the insulin and type 1 IGF receptors (Humbel, 1990), although IGF-II may stimulate some responses through its own receptor. For example, stimulation of glycogen

synthesis by IGF-II is not mediated by the type 2 IGF receptor, but through the type 1 or insulin receptors (Clemmons et al., 1992). In foetal growth and development, the actions of IGF-II may be mediated via the type 2 receptor. This is suggested by the observation of high levels of expression of the receptor in foetal tissues and a decline in postnatal receptor expression and circulating IGF-II levels (Sklar et al., 1989). Little is known about the responses mediated through the type 2 IGF receptor.

The hybrid type 1/insulin receptors can form in cells that express both receptor genes. They are composed of one insulin receptor α and β subunit and one type 1 receptor α and β subunit (Cohick and Clemmons, 1993). These hybrid receptors have ligand specificity profiles more comparable to the type 1 IGF receptor than the insulin receptor (Jones and Clemmons, 1995).

1.6 IGF Binding Proteins

1.6.1 Introduction

The insulin-like growth factor binding proteins (IGFBPs) are a family of six proteins that bind the IGFs with high affinity. These proteins were first identified by Hintz (1977) who reported the presence of specific somatomedin carrier proteins in serum. The IGFBPs exist extracellularly and, like IGFs, are synthesized in many tissues. Circulating and tissue IGFs are found tightly complexed to IGFBPs. There are at least six different IGFBPs which bind IGF-I and IGF-II with high specificity and affinity and no other peptide hormone is able to compete for binding. The IGFBP family possibly consists also of an additional four proteins that can associate with IGFs with lower affinity (Oh et al., 1997; Rechler, 1997). It has been reported that insulin is unable to bind to IGFBPs (De Vroede et al., 1985; Szabo et al., 1988), however, Heding et al. (1996), using a BIAcore instrument found human insulin to have a measurable, though extremely low, affinity for IGFBP-3. The BIAcore assay used in the study was found to be more sensitive for the detection of low affinity interactions than the competition assays used previously. The circulating IGF-IGFBP complex is biologically inactive (Rutanen and Pekonen, 1990).

The major functions of IGFBPs that are essential to coordinate and regulate the biological activities of the IGFs include; to act as transport proteins in plasma and to control the efflux of IGFs from the vascular space; to prolong the half-lives of the IGFs and regulate

their metabolic clearance; to provide a means of tissue and cell type specific localisation; and to modulate direct interaction of the IGFs with their receptors and thereby indirectly control biological actions (Jones and Clemmons, 1995). The IGFBPs are also believed to participate in the *in vivo* folding of IGFs (Hober et al., 1994), and inhibit IGF binding to the insulin receptor. The IGFBPs have also been shown to either inhibit or stimulate the growth-promoting effects the IGFs have on cell cultures, most likely by altering the interaction of IGFs with their cell surface receptors (Mohan et al., 1995). IGFBPs either inhibit or potentiate IGF activity depending on whether they are soluble or cell membrane associated (Bach and Rechler, 1995). Although some IGFBPs can limit access of IGF to receptors (Ritvos et al., 1988; Gopinath et al., 1989), IGFBP-3 and IGFBP-5 can associate with cell surfaces themselves (McCusker et al., 1991) and thereby, increase the amount of IGF bound to cell surfaces (McCusker et al., 1990), as well as inhibit the rate of receptor internalisation (Conover and Powell, 1991).

The IGFBPs, in turn, are regulated by the IGFBP proteases, a group of proteolytic enzymes that are capable of cleaving IGFBPs into smaller fragments with lower affinity for IGFs, thus enhancing IGF action. Various endocrine factors also play a role in IGFBP regulation (Levitt Katz et al., 1995). The affinity of IGFBPs for IGFs is also controlled by phosphorylation (Conover, 1995).

The growing number of IGFBP mutagenic studies described to date has focused on aspects of IGFBP biology such as extracellular matrix or heparin binding (Arai et al., 1996), integrin receptor binding (Jones et al., 1993), specific proteolysis (Conover, 1995; Chernausek et al., 1995) or phosphorylation (Jones et al., 1993) rather than systematic identification and characterisation of a common IGF binding motif.

Among the six binding proteins there is a high degree of sequence homology due to the conservation of the primary sequence of cysteine residues which are involved in disulfide bond formation. Human IGFBP-1, -2, -3, and -5 each contain 18 cysteine residues in conserved positions; IGFBP-4 contains two additional cysteines; and IGFBP-6 has only 16. The highest conservation is found in the N- (residues 1 to ~100) and C- (from residue 170 to the C-terminus) terminal cysteine-rich regions. Twelve conserved cysteines are found in the N-terminal domain and six in the C-terminal domain. The IGFBPs have a characteristic N-terminal cysteine pattern; GC(G/S)CCXXCAXXXXXXC, which can also be found in the human proteins MAC25 (Swisshelm et al., 1995) and connective tissue

growth factor. While all of the cysteine residues of IGFBPs are believed to participate in disulfide bonds (Sommer et al., 1993), it is still unknown whether all IGFBPs share the same disulfide-bridging pattern. The central, weakly conserved region (L-domain) contains most of the cleavage sites for specific proteases (Chernausek et al., 1995), phosphorylation and glycosylation sites and sites of association with other biomolecules such as heparin or the integrin receptor (Rechler, 1993; Jones and Clemmons, 1995).

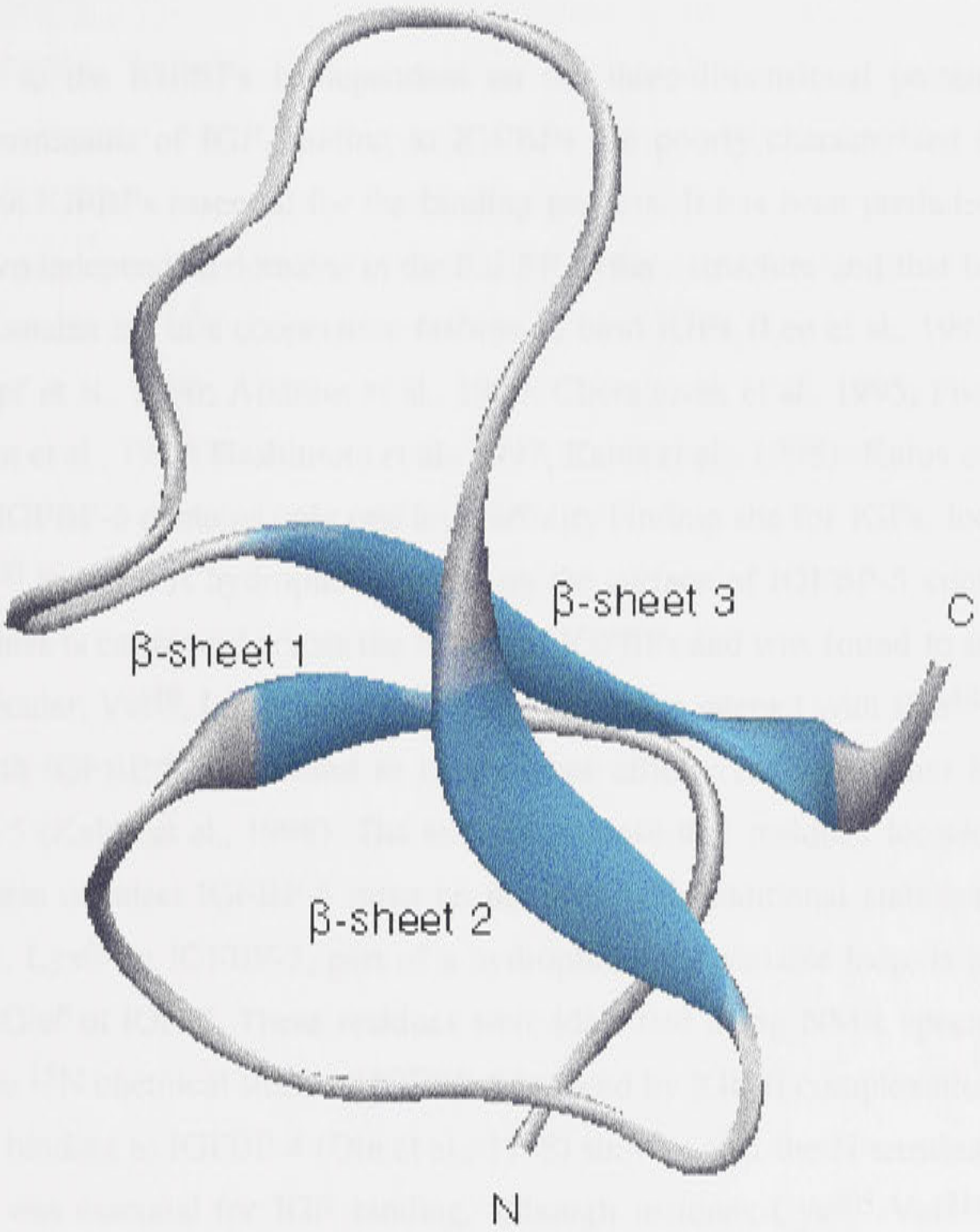

Figure 1.2: Ribbon drawing of mini-IGFBP-5 showing the three β-sheets (Kalus et al., 1998). The N- and C-termini of the fragment are labeled.

Disulfide mapping experiments and three-dimensional structures of IGFBPs are important for an understanding of the actions of these proteins. No high resolution structure is available for any intact member of the IGFBP family. The NMR structures of two N-

terminal fragments from IGFBP-5 which retain IGF-binding activity have recently been reported (Kalus et al., 1998), including a mini-IGFBP-5 which encompasses residues Ala^{40} to Ile^{92} of IGFBP-5 (Figure 1.2). The solution structure of mini-IGFBP-5 discloses a rigid and globular structure that consists of a centrally located three-stranded anti-parallel β-sheet, further stabilized by two inside packed disulfide bridges. Interestingly, no similar fold was found by the authors in database searches. A model for the IGF binding to IGFBP-3 has also been proposed (Spencer and Chan, 1995).

IGF binding to the IGFBPs is dependent on the three-dimensional protein structure, although determinants of IGF binding to IGFBPs are poorly characterized in terms of residues in the IGFBPs essential for the binding process. It has been predicted that there are at least two independent domains in the IGFBP tertiary structure and that both N- and C-terminal domains act in a cooperative fashion to bind IGFs (Lee et al., 1993; Bayne et al., 1988; Zapf et al., 1990; Andress et al., 1993; Chernausek et al., 1995; Fowlkes et al., 1995; Durham et al., 1997; Hashimoto et al., 1997; Kalus et al., 1998). Kalus et al. (1998) showed that IGFBP-5 contains only one high-affinity binding site for IGFs, located in the segment Ala^{40} to Ile^{92}. A hydrophobic patch on the surface of IGFBP-5 containing Leu and Val residues is conserved across the family of IGFBPs and was found to interact with IGFs. In particular, Val^{49}, Leu^{70} and Leu^{74} are proposed to interact with Phe^{48} of IGF-II. The full length IGFBP-5 was found to have higher affinity for IGF-I and IGF-II than mini-IGFBP-5 (Kalus et al., 1998). The authors propose that residues located in the C-terminal domain of intact IGFBP-5 must be involved with additional stabilisation of the IGF complex. Lys^{68} in IGFBP-5, part of a hydrophilic and variable loop, is proposed to interact with Glu^{6} of IGF-II. These residues were identified using NMR spectroscopy by changes in the ^{15}N chemical shifts of IGFBP-5 induced by IGF-II complexation. A recent study of IGF binding to IGFBP-4 (Qin et al., 1998) showed that the N-terminal sequence, Leu^{72}-Ser^{91}, was essential for IGF binding, although residues Cys^{205}-Val^{214} facilitated IGF binding. The regions of IGF-I involved in IGFBP binding and the conformation of the IGFBP may influence the distribution of IGF-I between receptors and IGFBPs at equilibrium (Clemmons et al., 1992).

The primary structure of IGFBP-2 contains six tyrosine residues: Tyr^{60} and Tyr^{71} are located in the cysteine rich N-terminal region, Tyr^{98} is located in the middle region, while Tyr^{213}, Tyr^{226} and Tyr^{269} lie in the cysteine-rich C-terminal region of the molecule (Upton et al., 1990). Hobba et al. (1996) used tyrosine iodination as a structural and

functional probe to establish that Tyr^{60} of IGF-I as essential for binding to bovine IGFBP-2. This method was used to minimise the structural disruption introduced by the random substitution of amino acids. Previously, this approach has been successfully used to identify tyrosine residues of IGF-I which are important for association with the type 1 IGF receptor (Maly et al., 1988; Schäffer et al., 1993) and tyrosine residues of both IGF-I and IGF-II which are important for association with IGFBP-2 (Moss et al., 1991). Tyrosine residues are found at the position equivalent to Tyr^{60} of IGFBP-2 in all IGFBP family members with the sole exception being IGFBP-1, where an alanine exists at this position (Lee et al., 1993). Adjacent residues, Val^{59}, Thr^{61}, Pro^{62} and Arg^{63} are well conserved in IGFBP family members, however, substitution of these residues did not reduce the IGF binding affinity of bovine IGFBP-2 (Hobba et al., 1998). Both the aromatic nature and the hydrogen bonding potential of the tyrosyl sidechain of Tyr^{60} were identified as important structural determinants of the IGF binding site of IGFBP-2 (Hobba et al., 1998). These researchers proposed that the hydroxyl group of Tyr^{60} participates in a hydrogen bond that is important for the initial complex formation with IGF-I and the stabilisation of this complex.

The high degree of structural similarity between IGF-I and IGF-II (Torres et al., 1995) suggests that both IGF molecules interact with IGFBPs through similar sidechain contacts. The N-terminal region of IGF-I has been implicated as an important region for binding to the IGFBPs. In particular, deletion of the first three residues of IGF-I leads to greatly reduced affinity for the binding protein (Ross et al., 1989; Bagley et al., 1989, King et al., 1992; and Francis et al., 1992). Long-[Arg^3]-IGF-I, which has the addition of a 13 amino acid N-terminal extension and the replacement of Glu^3 with Arg^3, has a seven-fold reduced affinity for IGFBP-3 (Heding et al., 1996). Clemmons et al. (1992) have found that mutation of the IGF-I residues 3 and 4 as well as 49-51 leads to reduced affinity for IGFBP-1. Terasawa et al. (1994) identified Phe^{26} of IGF-II as playing a role in changing the local structure of IGF-II to facilitate binding but it does not itself bind directly to IGFBPs.

Using NMR spectroscopy, Jansson et al. (1998) defined a structural binding epitope for IGF-I by comparing ^{15}N and ^{13}C chemical shifts in the presence and absence of IGFBP-1. A number of residues in IGF-I were affected by IGFBP-1 binding: Pro^2, Glu^3, Cys^6 and Gly^7, further confirming the importance of the N-terminus for IGFBP binding. Other residues affected were Pro^{28}-Gly^{30}, Gly^{32}, Arg^{36}, Arg^{37}, Gln^{40}-Gly^{42}, Pro^{63}, Lys^{65},

Pro^{66} and Lys^{68}-Ala^{70}. This area spans most of one face of the molecular surface of IGF-I. Three arginine sidechains: Arg^{36}, Arg^{37} and Arg^{50} were shown to be involved in IGFBP-1 binding. Arg^{36} and Arg^{37}, together with Arg^{21} and Arg^{56} also participate in binding to the type 1 IGF receptor. Jansson et al. (1998) found that the IGF-I/IGFBP-1 complex does not associate with the type 1 IGF receptor. This confirms that only free IGF-I is capable of binding to the type 1 IGF receptor, since the binding determinants on IGF-I for IGFBP-1 and the type 1 IGF-I receptor overlap. The binding of IGFBPs and the type 1 IGF receptor to IGF-I is competitive, since the presence of either will partially block the other's binding surface on IGF-I. This competitive binding of IGFBPs and IGF receptors for IGFs is supported by Kalus et al (1998) who found that IGFBP-5 inhibits IGF-stimulated autophosphorylation of the receptor and complete inhibition is achieved when IGFBP-5 is in excess of the IGFs.

The structures of each of the IGFBPs are discussed briefly below along with their roles in circulation. Some IGFBPs inhibit IGF actions while others enhance the mitogenic effects of IGFs. Levitt Katz et al. (1995) have provided an extensive review of the individual IGFBPs including their physiological characteristics and regulation and this review provides the basis for the information presented below.

1.6.2 IGFBP-I

IGFBP-1 is a 25 kDa protein that is able to inhibit or enhance IGF action *in vitro*. Its secretion by the liver into serum is regulated by insulin while glucocorticoids, glucagon, cAMP and somatostatin have a stimulatory effect on IGFBP-1 production. IGFBP-1 serum levels are also dependent on age and food intake. Its main biological activity appears to be in carbohydrate regulation. IGFBP-1 carries IGF-I in the circulation as a binary complex (Twigg and Baxter, 1998). The method of cellular binding is via an Arg-Gly-Asp (RGD) sequence near the carboxyl terminal. Glu^{3} and Phe^{49} of IGF-I were identified by Dubaquié and Lowman (1999) as major specificity determinants for IGFBP-1. The corresponding alanine mutations, $Glu^{3}Ala$ and $Phe^{49}Ala$, selectively decreased IGFBP-1 binding by 34- and 100-fold respectively. Clemmons et al. (1992) also showed that optimal binding of IGF-I and IGF-II to IGFBP-1 requires contacts between residues Glu^{3} and Thr^{4} at the N-terminus and Phe^{49}-Ser^{51} in the A-domain. The A-domain residues 48-50 and Phe^{26} of IGF-II are important for binding to IGFBP-1 (Bach et al. 1993).

1.6.3 IGFBP-2

IGFBP-2 is a 31 kDa protein which is found in seminal plasma in high concentrations and in cerebrospinal fluid and serum in lower concentrations. It is secreted by many cell types and is expressed in many foetal and adult tissues. It has an inhibitory effect on IGF actions. IGFBP-2 carries IGF-I in the circulation as a binary complex (Twigg et al., 1998). IGFBP-2 has a preference for IGF-II over IGF-I (Jones and Clemmons, 1995). The method of cellular binding is through the RGD sequence described above.

Hobba et al. (1998) proposed that Tyr^{60} is located in the IGF binding interface of IGFBP-2 and that modification or mutagenic replacement of Tyr^{60} directly disrupts contacts between IGF and IGFBP-2. However, the alternative possibility that Tyr^{60} modification and mutagenic replacement indirectly affect IGF binding for example, by preventing a change in conformation that is necessary for high affinity IGF binding, was not eliminated by the authors. The presence of an aromatic residue at position 60 of IGFBP-2 reduces the rate of formation of the IGFBP-2/IGF complex, but also enhances the stabilisation of the complex. Hobba et al. (1996) concluded that both the aromatic nature and the hydrogen bonding potential of the tyrosyl sidechain of Tyr^{60} are important structural determinants of the IGF-binding site of IGFBP-2. Hobba et al. (1998) found that Tyr^{60} mutants had disproportionate sensitivity toward IGF-I binding, suggesting that the IGFBP-2/IGF-II complex contains additional points of molecular interaction that are absent in the IGFBP-2/IGF-I complex. Forbes et al. (1998) found that the C-terminal cysteine rich domain of IGFBP-2 contains determinants of IGF-II binding specificity.

Clemmons et al. (1992) showed that substitutions of residues Phe^{49}-Ser^{51} of IGF-I result in the greatest loss of binding activity for IGFBP-2, while substitutions of Glu^{3} and Thr^{4} of the B-domain result in marked attenuation but not total loss of binding affinity for IGFBP-2. The A-domain residues 48-50 of IGF-II are important for binding to IGFBP-2 (Bach et al., 1993).

1.6.4 IGFBP-3

The majority of IGF-I present in serum circulates as a complex with IGFBP-3, and another protein, acid-labile subunit (ALS) (Baxter et al., 1989). IGFBP-3 is synthesized in many cells and is the major IGFBP in postnatal serum. Levels of IGFBP-3 rise throughout childhood and further increase during puberty, remaining stable throughout

most of adult life and slowly falling during aging. In plasma, IGFBP-3 circulates as a glycosylated protein of 40-44 kDa and when IGF binds is found as part of a 150 kDa complex that includes ALS. Proteolytic cleavage of the IGFBP-3 150 kDa complex reduces its affinity for IGF and the functional properties of IGFBP-3 are altered such that IGF is released and becomes available to target cells. The primary regulators of IGFBP-3 are GH and proteases. *In vitro*, the action of IGFBP-3 has been described to be IGF inhibitory, IGF stimulatory or IGF independent, depending on the experimental system used (reviewed in Levitt Katz et al., 1995). Heding et al. (1996) found non-glycosylated IGFBP-3 to have a comparable affinity for IGF-I and IGF-II with a slight preference for IGF-II.

No IGF-I sidechain specificity determinant was found for IGFBP-3 by alanine screening (Dubaquié and Lowman, 1999). Instead, their results suggest that the N-terminal backbone region of IGF-I is important for binding to IGFBP-3. The A-domain residues Phe^{48}, Arg^{49}, and Ser^{50} of IGF-II are important for binding to IGFBP-3 (Bach et al., 1993).

1.6.5 IGFBP-4

IGFBP-4 carries IGF-I in the circulation as a binary complex (Twigg et al., 1998). IGFBP-4 is a 24 kDa protein which also appears in a glycosylated 28 kDa form. It is found in serum and seminal plasma as well as in numerous cell types including prostate and bone. Cellular binding is through membrane association. IGFBP-4 is inhibitory to IGF action. Fowlkes and Freemark (1992) demonstrated that IGF-I and IGF-II can directly promote the proteolytic degradation of IGFBP-4 into fragments that do not bind IGFs, providing a novel mechanism by which the IGFs may increase their own availability and/or activity in biological fluids.

Clemmons et al. (1992) found any alteration of the B-domain amino terminus or helix 1 of IGF-I results in a loss of affinity for IGFBP-4. Substitution of Phe^{49}-Ser^{51} of IGF-I also results in reduced affinity for IGFBP-4. Helices 1, 2 and 3 of IGF-I are all necessary for optimum binding to IGFBP-4, while IGFBP-4 has additional determinants in the D-domain of IGF-I and the carboxyl terminus of the B-domain. Clemmons et al. (1992) concluded that the binding pocket of IGFBP-4 is larger than the binding pockets of IGFBP-1, 2, and 3 and suggests that it may assume a different conformation to IGFBP-1, 2, and 3. The A-domain residues Phe^{48}, Arg^{49}, and Arg^{50} of IGF-II are important for

binding to IGFBP-4 (Bach et al., 1993).

1.6.6 IGFBP-5

IGFBP-5 is a glycosylated 29 kDa protein which has been observed in cerebrospinal fluid, serum and rapidly growing foetal tissues. IGFBP-5 potentiates the action of IGF-I on smooth muscle cells, fibroblasts or osteoblasts (Bautista et al., 1991). IGF-I is carried by IGFBP-5 in a ternary complex with ALS (Twigg et al., 1998). IGFBP-5 binds to the extracellular matrix of heparins and this binding leads to a 8- to 17-fold decrease of the IGFBP-5 affinity for IGF-I and enhances the biological action of IGF-I on the cells (Jones et al., 1993; Arai et al., 1996; Campbell and Andress, 1997). IGFBP-5 has been shown to have an inhibitory effect on the growth of carcinoma C33A cells (Higo et al., 1997) and on the induction of alkaline phosphatase in foetal rat calvarial cells by OP-1 (Yeh et al., 1997), demonstrating the ability of IGFBP-5 to act as an inhibitor of cell proliferation or differentiation. IGF-I receptor binding assays of IGFBP-5 showed that IGFBP-5 inhibits the binding of IGFs to the type 1 IGF receptor, resulting in reduction of receptor stimulation and autophosphorylation and the inhibition of IGF biological activity (Kalus et al., 1998).

IGFBP-5 has been shown to contain one high affinity binding site for IGFs, located in the N-terminal segment Ala^{40} to Ile^{92} (Kalus et al., 1998). The NMR solution structure (Figure 1.2) of this segment demonstrated a globular structure consisting of a centrally located three-stranded anti-parallel β-sheet (Kalus et al., 1998). Kalus et al. (1998) found that the binding of IGF-II to its receptor is sterically hindered by the C-terminal domain of IGFBP-5. Clemmons et al. (1992) found IGFBP-5 to be the binding protein with the most stringent requirements for IGF binding. Mutations in the A, B and D-domains of IGF-I result in reduced affinity for IGFBP-5 suggesting a larger binding pocket than IGFBP-1, 2 or 3. In contrast to IGFBP-4, alteration of the N-terminal region of IGF-I greatly reduces the affinity for IGFBP-5 which indicates that IGFBP-5 has a more stringent requirement for contact at Glu^{3} and Thr^{4} of IGF-I relative to IGFBP-4. As for other IGFBPs, the A-domain residues Phe^{48}, Arg^{49} and Ser^{50} of IGF-II are important for IGFBP-5 binding (Bach et al., 1993).

1.6.7 IGFBP-6

IGFBP-6 carries IGF-I in the circulation as binary complex (Twigg et al. 1998). IGFBP-6

is a glycosylated 34 kDa protein found in cerebrospinal fluid and serum. It is the only IGFBP that is specific for one of the IGFs, whereby it has higher specificity for IGF-II over IGF-I (Twigg et al., 1998). Substitution of the A-domain residues Phe^{48}, Arg^{49}, and Ser^{50} of IGF-II resulted in more than 50-fold reduced binding to IGFBP-6. Phe^{26} of IGF-II is also important for binding to IGFBP-6 (Bach et al., 1993).

1.7 Biological Activity

The growth in research involving the IGF system is to a large extent due to findings indicating a pivotal role for these proteins in cancer and several metabolic disorders such as neurodegenerative diseases and osteoporosis (Walsh, 1995; Rosen et al., 1997; Werner and De Roth, 1998). The original somatomedin hypothesis of Daughaday et al. (1972) stated that GH acts on peripheral tissues, not directly, but by stimulating the liver synthesis of IGF, which then mediates the somatogenic actions of GH in the target tissues in an endocrine manner. However, the findings that IGFs are produced by many tissues in addition to the liver, and that this production is also GH stimulated, have changed the concept of IGF action from primarily hormonal to local, i.e. paracrine or autocrine (Humbel, 1990).

1.7.1 In Vitro Effects

The *in vitro* effects of IGFs are classified into short-term insulin-like metabolic effects and long-term mitogenic effects. The rapid effects are readily observed with classical insulin target cells and, in terms of the cell response, are quantitatively indistinguishable from those of insulin. Metabolic effects include stimulation of glucose uptake and glycogen and lipid synthesis in adipose tissue. In certain cell types, short-term effects may be mediated via the insulin receptor, in other cell types, via the type 1 IGF receptor (Walton et al., 1990).

The long-term effects are mostly mediated by the type 1 IGF receptor and include stimulation of mitosis and cell differentiation, enhanced synthesis of DNA, RNA, protein, glycogen, lipid, and steroid hormones and the inhibition of protein breakdown (Humbel, 1990).

A negative feedback loop also exists in which high IGF-I and IGF-II levels inhibit GH synthesis and secretion in pituitary cells. IGFs also stimulate transport, including calcium uptake and transferrin receptor up regulation. It is believed that the type 2 receptor does not usually mediate any of the above-mentioned effects (Walton et al., 1990).

In vitro, IGF-II is an autocrine growth factor in neuroblastoma and rhabdomyosarcoma cell lines and an autocrine differential factor for skeletal muscle (Bach et al., 1993). IGF-II can stimulate a response through its own receptor, including the stimulation of DNA synthesis in a human erythroleukemia cell line and a rat cell line, proliferation of K-562 cells (Tally et al., 1987), glycogen synthesis in hepatoma cells (Hari et al., 1987), alkalinisation of cells of the proximal tubular segment of canine kidney by stimulating Na^+/H^+ exchange across the brush border membrane (Mellas et al., 1986) and stimulation of Ca^{2+} influx and DNA synthesis in competent Balb/c 3T3 cells primed with epidermal growth factor (Lobel et al., 1988). Minniti et al. (1992) found that in rhabdomyosarcoma cells, IGF-II stimulates two different responses mediated by distinct receptors: 1) a mitogenic response through the type 1 IGF receptor and 2) a motility response through the type 2 IGF receptor.

1.7.2 In Vivo Effects

In vivo studies with IGFs have confirmed that these peptides have both an acute insulin-like action as well as a chronic growth promoting effect and the effects of IGFs depend on their mode of administration, with intravenous bolus injections causing acute insulin-like effects, while long-term subcutaneous administration induces growth (reviewed in Zapf, 1995). The acute metabolic actions of IGF-I and IGF-II were observed by Zapf et al. (1979) (reviewed in Zapf, 1995) when a bolus 40 µg of recombinant human IGF-I was injected into rats intravenously, and a resultant transient hypoglycaemia with a maximum at 15 min after injection was observed (reviewed in Zapf, 1995). Even more pronounced was the fall in blood sugar when a bolus of 100 µg/kg body mass was injected into healthy human volunteers. Acute injections temporarily result in such high concentrations of IGF-I that they exceed the binding capacity of the IGFBPs (reviewed in Zapf, 1995). Bolus injection of IGF-II also induced hypoglycaemic effects with decreased potency (reviewed in Zapf, 1995).

Subcutaneous infusion of IGF-I was administered over six days to hypophysectomised rats (in which the pituitary gland has been removed) (Guler et al., 1989). This resulted in increased body weight, long bone growth and accumulation of lean body mass. Similar results were obtained for normal rats and Snell dwarf mice. Body weight gain, tibial epiphyseal width and epiphyseal growth rate were stimulated to the same extent by either GH or IGF-I. The accumulated longitudinal bone length was slightly more following administration of GH than of IGF-I. GH increased the weights of two examined muscles in comparison to IGF-I, although IGF-I and not GH increased the weights of the kidney, spleen and thymus, and decreased the weight of the fat pads. Due to the slow, continuous infusion, no hypoglycaemia occurred. IGF-II was found to be less potent than IGF-I in promoting growth (Guler et al., 1989). Normal growth of diabetic rats was restored by infusion of recombinant human IGF-I without normalisation of the blood sugar level and insulin acts via an increase in IGF-I synthesis on growth of diabetic rats (Scheiwiller et al., 1986).

Chronic infusion of exogenous IGF-I has a number of other effects in addition to stimulating growth. Recombinant IGF-I infused into healthy human subjects resulted in inhibition of insulin secretion, increases in renal plasma flow and glomerular filtration rate (Guler et al., 1989). In comparable studies with rodents, the kidney weight was increased substantially (Kurtz et al., 1988). The same research group found that infusion of exogenous IGF-I stimulated erythrocyte synthesis and proposed that this effect may have been due to erythropoeiten synthesis. In both starved mice and diabetic rats, IGF-I reduces body weight losses. Spencer et al. (1988) found that IGF-I directly applied to wounds caused an increased growth of fibrous tissue in rats and rabbits. Further study is required to assess the efficacy of IGF-I as a therapy for impaired renal function, reduced erythrocyte synthesis, or wound healing.

DeChiara et al. (1990) examined the developmental role of IGF-II by introducing mutations at the IGF-II gene locus in the mouse germ line. The authors demonstrated the first direct evidence for a physiological role of IGF-II in embryonic growth, since growth deficient animals which were otherwise normal and fertile resulted from the inactivated IGF-II gene. The growth-deficiency phenotype becomes apparent at least as early as embryonic day 16 and persists after birth. The possibility exists that the growth-deficiency phenotype is the indirect consequence of impaired placental trophic functions. IGF-II stimulates a number of responses through its own receptor including stimulation of

calcium ion influx in 3T3 fibroblasts (Nishimoto et al., 1987) and amino acid uptake in human myoblasts (Shimizu et al., 1986).

A summary of the biological actions and interactions of IGFs with their receptors and binding proteins is presented in Table 1.2.

Table 1.2. Biological actions of IGFs *in vivo* and *in vitro*

In Vivo	*In Vitro*
Glucose Metabolism	**Stimulate transport**
Increased glucose uptake	Nutrient uptake (glucose, amino acids)
Decreased hepatic glucose production	Calcium influx
Increased apparent insulin sensitivity	Transferrin receptor upregulation
Hypoglycaemia	**Stimulate Biosynthesis**
Fat Metabolism	DNA synthesis
Decreased serum ketones	RNA synthesis
Decreased serum free fatty acids	Protein synthesis
Decreased triglycerides	Glycogen synthesis
Protein Anabolism	Lipid synthesis
Increased protein synthesis	Steroid synthesis
Decreased nitrogen secretion	**Inhibit Catabolism**
Increased total body protein accretion	Protein breakdown
Increased body organ weight	**Stimulate Cell Division**
Improved wound healing	of cartilage, muscle and bone cells
Reduces weight loss in starvation	**Stimulate Cell Differentiation**
Renal Function	**Feedback Loop**
Increased glomerular filtration rate	Inhibits GH release from pituitary cells
Increased renal plasma flow	
Stimulates kidney growth	
Increased speed of recovery from ischaemic acute renal failure	
Hormones	
Decreased growth hormone	
Decreased glucagon	
Increased catecholamines	
Growth	
Increase in hypopituitary and GH-insensitive subjects	

1.8 Insulin-Like Growth Factor Analogues

A number of analogues of both IGF-I and IGF-II have been produced. These include IGFs with various site mutations (e.g. Sakano et al., 1991), deletions (e.g. King et al., 1992; De Wolf et al., 1996) and fusion proteins (e.g. Francis et al., 1992). The majority of these analogues were produced using recombinant gene technology in order to investigate the binding sites of IGF-I and IGF-II to either the IGFBPs or the type 1, type 2 or insulin receptors.

1.8.1 N-terminal Modifications

Some of these analogues have been found to be more potent than the respective parent IGF-I or IGF-II. For example, a naturally occurring analogue, Des-(1-3)-IGF-I, which has a sequence identical to IGF-I but with the N-terminal tripeptide Gly^1-Pro^2-Glu^3 omitted, was found to be more potent at stimulating protein or DNA synthesis and inhibiting protein breakdown (Francis et al., 1986; Francis et al., 1988; and Ballard et al., 1987). Ross et al. (1989) concluded that this enhanced potency of Des-(1-3)-IGF-I is a consequence of its poor binding to IGFBPs. In order to further investigate the role of the N-terminus of IGF-I with respect to its biological properties, analogues of IGF-I truncated by 1-5 amino acid residues from the N-terminus were chemically synthesized. (Bagley et al., 1989). Increased biological potency was observed for Des-(1-3)-IGF-I and Des-(1-4)-IGF-I presumably due to a lower affinity for binding to the IGFBPs. However, Des-(1-4)-IGF-I has a lower potency than Des-(1-3)-IGF-I as a consequence of a lower affinity for the type 1 IGF receptor.

The same research group found [Gly^3]-IGF-I and [Arg^3]-IGF-I to be more potent than IGF-I in biological assays measuring stimulation of protein synthesis and DNA synthesis or inhibition of protein breakdown (King et al., 1992). Again, both analogues bound poorly to IGFBP-2 and slightly less well to the type 1 IGF receptor. It would seem from these results that the binding of IGF-I to IGFBP-2 is mediated, at least in part, by electrostatic interactions involving Glu^3 of IGF-I. It was concluded that reduced binding to IGFBPs rather than increased receptor binding is responsible for the greater biological

potency of the analogues compared with IGF-I. Following this, Francis et al. (1992) found several fusion peptide analogues to be more potent than the native IGFs in stimulating protein and DNA synthesis and inhibiting protein breakdown. The analogues comprise the first 11 amino acids of methionyl porcine growth hormone (pGH) and the dipeptide Val-Asn, followed by the native IGF-I sequence and were named Long-IGF-I. Two of the fusion peptides also contain substitutions at position 3 with Gly and Arg replacing Glu. The order of biological potency was found to be Long-[Arg^3]-IGF-I > Des-(1-3)-IGF-I > Long-[Gly^3]-IGF-I > Long-IGF-I > IGF-I. Investigations of receptor and IGFBP association by these IGF-I analogues with N-terminal modifications show increased biological potency, which was attributed to a decreased affinity for the IGFBPs (e.g. Francis et al., 1992). The Long-[Arg^3]-IGF-I species is the most potent of these analogues because it has the weakest IGFBP affinity. The N-terminal IGF-I mutants Des(1-3)-IGF-I and Long-[Arg^3]-IGF-I both had reduced affinity for IGFBP-3. Long-[Arg^3]-IGF-I has much lower affinity than Des(1-3)-IGF-I for IGFBP-3, indicating that the 13 amino acid N-terminal extension and the sidechain charge reversal at residue 3 of Long-[Arg^3]-IGF-I impedes IGFBP-3 binding.

1.8.2 A- and B-Domain Mutants

Mutations of IGF-I in the B- (residues 1-29) and A- (residues 42-62) domains were produced with the aim of identifying IGF-I residues that contribute to the specificity of binding to the type 1 IGF receptor as opposed to the insulin receptor (Shooter et al., 1996). The B- and A- domains of IGF-I share a high degree of sequence and structural similarity to the B- and A-chains of insulin (Blundell et al., 1978). Amino acid residues not conserved between insulin and IGF-I were target sites for altering receptor binding specificity. Ala^8, Gln^{15} and Ala^{62} of IGF-I were substituted because of their close proximity to the proposed receptor binding region (Cooke et al., 1991). Substitutions were intended to decrease binding to the insulin receptor by maximising differences in amino acid sidechain character between IGF-I and insulin at each position. In addition, a Thr^4His analogue was chosen to evaluate the role of the N-terminal region in receptor binding, and a Met^{59}Phe mutation was made because Met^{59} is close in space to Tyr^{24} and because its sidechain resides in the cleft between the C- and D-domains.

The altered specificity of Phe^{59}-IGF-I favoured the type 1 IGF receptor over the insulin receptor, which may be due to steric effects produced by reduced flexibility in the C- and

D-domains although the direct involvement of this residue in defining receptor binding affinity could not be ruled out. Relative to IGF-I, [Leu^{8}]-IGF-I had significantly lower affinity for the insulin receptor than it has for the type 1 IGF receptor. Substitution at position 62 had minimal impact on insulin receptor binding but reduced affinity for the type 1 IGF receptor. The substitution at position 4 resulted in an increase in insulin receptor binding with no significant change in relative affinity for the type 1 IGF receptor. Substituting Ser for Gln at position 15 had no substantial effect on receptor binding or other biological properties. It was concluded that the co-evolution of the IGF and insulin receptor/ligand systems has resulted in subtle differences in the A- and B- regions of each ligand, and that these two regions are important for defining receptor binding specificity (Shooter et al., 1996).

1.8.3 C-Domain Mutants

Recently the solution structure of a mini-IGF-I has been determined by NMR spectroscopy (Wolf et al., 1996). This inactive mutant lacks the C- domain, which is in extended conformation in IGF-I. Deletion of the C- domain of IGF-I results in an altered relative orientation of the three helical segments (within the A and B domains) with a reorganisation of the hydrophobic core and displacement of the Phe^{23}-Tyr^{24}-Phe^{25}-Asn^{26} segment which is known to be important for receptor binding. It has been shown by Gill et al. (1996) that this mutant has much lower receptor binding affinity and loss of biological activity. While mini-IGF-I is still recognised by IGFBP-3, it binds with an association constant one sixth of that for wild-type IGF-I.

Heding et al. (1996) investigated the binding kinetics of IGF-I analogues toward IGFBP-3. 4-Gly IGF-I, which has the C-domain replaced by a 4-Gly bridge, has a 3-fold lower affinity for IGF-I. The two-fold lower affinity of mini-IGF-I for IGFBP-3 compared with that of 4-Gly IGF-I is attributed to the complete absence of the C-domain in mini-IGF-I, since this presumably causes a greater conformational change in the molecule than the replacement of the C-domain residues with a 4-Gly bridge. The relatively high affinity of these analogues for IGFBP-3 is in contrast to their very low affinity for the type 1 IGF receptor (Gill et al., 1996; Bayne et al., 1989; Cascieri et al., 1989). Three IGF-I analogues in which the disulfide bonds have been swapped, IGF-I swap, 4-Gly IGF-I swap and mini-IGF-I swap all have low affinities for IGFBP-3, indicating that swapping the disulfide bridges in IGF-I greatly impairs the hormone's ability to bind to IGFBP-3 (Heding et al.,

1996). The disulfide swapped IGF-I analogues also have extremely low receptor affinity (Gill et al., 1996), which indicates that swapping of the disulfide bonds causes a major conformational change of the entire protein as found by Miller et al. (1993).

Cara et al. (1990) prepared a two-chain insulin/IGF-I hybrid that contains a synthetic peptide related to residues 22-41 of IGF-I linked via peptide bond to Arg^{B22} of des-octapeptide-(B23-B30)-insulin and applied the analogue to the analysis of ligand interactions with the type 1 IGF and insulin receptors of placental plasma membranes. The major aspects of structure relevant to the conferral of receptor binding affinity lie in the C-terminal region of the insulin B chain and in the C-terminal region of the IGF-I B-domain and in its C-domain.

1.8.4 Indirect Agonists

Recently, a class of IGF-I analogues have been prepared which have normal binding affinities for IGFBPs but reduced binding to the type 1 IGF receptor and the insulin receptor (Milner et al., submitted for publication). These analogues which involve mutation of the tyrosine residues at positions 24, 31 and 60 to either alanine or leucine, increase the levels of endogenous IGF-I by selective displacement of IGF-I from its association with IGFBPs, thus promoting biological activity through action at the IGF receptor. This class of compounds are referred to as 'indirect agonists' (Milner et al., submitted for publication). These mutations were based on the observation by Bayne et al. (1990) that the three tyrosine residues at positions 24, 31 and 60 in the IGF-I molecule are involved in high affinity binding of IGF-I to the type 1 receptor.

The IGF-I analogues prepared by Milner et al. (submitted for publication) included $[Leu^{24}]$-IGF-I, $[Ala^{31}]$-IGF-I, $[Leu^{60}]$-IGF-I, $[Leu^{24}Ala^{31}]$-IGF-I and $[Ala^{31}Leu^{60}]$-IGF-I. $[Leu^{60}]$-IGF-I was chosen for further study as it exhibits poor binding to the type 1 IGF receptor and normal binding to acid stripped IGFBPs, weak stimulation of protein synthesis and good binding to ternary and binary complexes in human serum. To demonstrate the *in vitro* effect of displacement of IGF-I from binding proteins by an 'indirect agonist', $[Leu^{60}]$-IGF-I was added to a mixture of IGF-I and bovine IGFBP-2 (bIGFBP-2) in the presence of rat myoblast monolayers. When $[Leu^{60}]$-IGF-I was added to the cells with bIGFBP-2, there was very little stimulation of protein synthesis, even in the absence of inhibitory IGFBPs. When $[Leu^{60}]$-IGF-I was added to the cells in the

presence of a stoichiometric concentration of IGF-I, endogenous IGF-I was displaced from IGFBP-2, and more IGF-I was available to stimulate the receptor.

1.8.5 Disulfide Intermediates

Analogues have also been designed to investigate the oxidative refolding pathway of IGF-I (Narhi et al., 1993). The oxidative folding of human IGF-I yields two products of similar thermodynamic stability (Raschdorf et al., 1988; Meng et al., 1989; Hober et al., 1992; Miller et al., 1993). Analogous isomers of human proinsulin or IGF-II were not seen upon oxidative refolding. Miller et al. (1993) monitored the kinetics of the IGF-I folding pathway by high performance liquid chromatography (HPLC). Two disulfide isomers were obtained as products, which differ by interchange of a strained disulfide bond: one with native insulin-like pairing [6-48; 18-61; 47-52] (designated native IGF-I; 60% yield) and the other with alternative pairing [6-47; 18-61; 48-52] (designated IGF-swap; 40% yield). The predominant early folding intermediate contained the single disulfide 18-61, which is common to both products. Other folding intermediates include [6-48; 18-61] which is an immediate precursor of native IGF-I and [18-61; 6-47], an immediate precursor to IGF-swap with non native disulfide 6-47.

Spectroscopic studies employing circular dichroism (CD), infrared spectroscopy (FTIR), two dimensional ^{1}H-NMR, and photochemical dynamic nuclear polarisation (photo-CIDNP) indicated that IGF-I and IGF-swap adopt similar secondary structures but distinct tertiary folds (Miller et al., 1993). It is clear that a more general model of protein folding must include contributions from accessory and regulatory proteins, including chaperones and enzymes such as the disulfide and proline isomerases (Krebs et al., 1983; Jaenicke, 1991; Martin et al., 1991; Gething & Sambrook, 1992). Protein folding may also be influenced by specific contributions from proline sequences (Zhu et al., 1989; Winter & Sorensen, 1991; Weissman & Kim, 1992) and by differences in the kinetics of alternative pathways (Baker et al., 1992).

The role of the three native disulfide pairings in the folding and function of IGF-I was also reported (Narhi et al., 1993). Analogues containing pairwise Cys to Ser or Cys to Ala substitutions were expressed in *E. coli*, purified and analysed with respect to receptor binding, solution structure and thermodynamic stability. The analogue lacking all three disulfide bonds (designated des-Cys-IGF-I) was found to be inactive and unfolded.

Introduction of the [18-61] disulfide bond, previously shown to occur in an early intermediate in oxidative refolding (Miller et al., 1993), resulted in a compact folded state with low but significant biological activity. Additional, but incomplete structural organisation and biological activity were observed following the introduction of either the [6-48] or the [47-52] disulfide bonds. Successive formation of native disulfide bonds leads to stepwise stabilisation of native structural elements. The one- and two- disulfide analogues may be viewed as molten globules, i.e. compact partially folded states containing defined elements of secondary structure with fluctuating tertiary orientation (Baum et al., 1989; Kuwajima et al., 1991).

The folding pathway of Long-[Arg^3]-IGF-I has been compared with that of IGF-I by Milner et al. (1999). Unlike IGF-I, which yields 45% of the native structure and 24% of the alternative structure ([18-61], [6-47] and [48-52] disulfide bonds), Long-[Arg^3]-IGF-I yields 85% and 10% of these respective forms. Three native-like intermediates were identified, that appear to have a major role in the *in vitro* refolding pathway of Long-[Arg^3]-IGF-I; a single [18-61] disulfide intermediate, an intermediate with [18-61] and [6-48] disulfide bonds, and another with [18-61] and [47-52] disulfide bonds. Milner et al. (1999) proposed that the [18-61], [6-48] intermediate forms the native structure, not by direct formation of the [47-52] disulfide bond, but by rearrangement via the [18-61] intermediate and a productive [18-61], [47-52] intermediate. In this pathway, the last disulfide bond to form involves Cys^6 and Cys^{48}. This pathway was applied to IGF-I and it was suggested that the divergence in the *in vitro* folding pathway of IGF-I is caused by non-native interactions involving Glu^3 that stabilize the alternative structure. The interactions that stabilize the "off-pathway" intermediates are reduced for Long-[Arg^3]-IGF-I due to the N-terminal extension and the charge reversal at position 3.

1.9 NMR Approach to Protein Structure Determination

1.9.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy can give detailed information about protein structure, dynamics and interactions. NMR in bulk matter was demonstrated independently by Bloch et al. (1946) and Purcell et al. (1946). Shortly after, it was discovered that although the signals were coming from the nucleus, the nuclear resonant frequency was sensitive to the chemical environment (Proctor and Yu, 1950). This

phenomenon is known as chemical shift. With improvements in the homogeneity of magnets, the phenomenon of spin-spin coupling arising from interactions between nuclei sharing bonding electrons was discovered (Gutowsky et al., 1951). In 1955, Overhauser suggested that the intensity of an NMR signal would be enhanced by irradiation of unpaired electron spins in a sample. This lead to the nuclear Overhauser experiment where the mutual dipolar interaction of two protons close in space can be determined, thus giving distance information. This experiment is exploited in protein structure determination. Two important technical advances were the introduction in the late 1960s of pulse methods with Fourier transformation (Ernst et al., 1966) and, a decade later, two-dimensional techniques (Aue et al., 1976).

In the late 70s and early 80s, the role of NMR in structural biology became established. The first two dimensional spectrum of a protein was obtained by Nagayama et al. (1977). Kurt Wüthrich and colleagues eg., Wüthrich et al., 1982 demonstrated that protein conformation can be calculated in a systematic way provided that the ^{1}H resonances can be assigned to particular chemical groups in the protein. The method by which this is done is termed the sequential assignment procedure and is described in section 1.9.2. The first polypeptide structure determined from NMR data was derived by Williamson et al. (1985).

The information content available from the sequential assignment method depends on the molecular mass of the protein since, with increasing mass, the linewidths become broader and the overlap of resonances becomes more severe, meaning resolution and assignment become increasingly difficult (Campbell et al., 1991). This problem of overlap is alleviated by the use of isotopic labels together with multidimensional pulse techniques.

Tremendous growth has taken place in the field of NMR spectroscopy and its application to the study of macromolecules. NMR can not only be considered an alternative to x-ray crystallography when protein crystals are not available, but also a complementary technique that provides an alternative perspective to the protein structure and dynamics. Higher field magnets have been constructed, with spectrometers operating at ^{1}H frequencies as high as 800 MHz now available, resulting in improved sensitivity and resolution. The development of multinuclear, multidimensional NMR experiments based on sophisticated pulse schemes, coupled with either specific or uniform isotopic labelling of the molecule, has greatly enhanced the utility of NMR for the study of molecular

structure and dynamics (Kay and Gardner, 1997).

The development of probes with actively shielded gradient coils has made pulsed field gradient technology feasible in high resolution spectroscopy, resulting in substantial enhancements in spectral quality, via an improvement in artefact and solvent suppression (Kay, 1995). Before the development of gradient technology, artefacts in NMR experiments were eliminated by phase cycling, i.e. by repeating the experiment several times with different phases for the RF pulses and/or the receiver. The phase cycling is devised to ensure that the desired coherence transfer pathway is retained and unwanted coherence transfer pathways cancel at the end of the phase cycle. This process requires excellent spectrometer stability, often for a period of days. In practice, data acquisition times often limit the utility of phase cycling. The use of pulsed field gradients permits the suppression of artefacts on a per scan basis, thereby reducing the need for long term spectrometer stability and permitting acquisitions to be set according to the desired resolution and signal-to-noise (Hurd and John, 1991; Vuister et al., 1992). Pulsed field gradients are introduced into the pulse sequence to refocus the coherences in the desired pathway and dephase coherences that follow unwanted coherence pathways.

The solution of a protein structure by NMR can best be summarized as a method of translating distance and dihedral angle information into atomic coordinates. This process follows several well established steps. The initial step is the sequence-specific assignment of NMR resonances (Wüthrich et al., 1986). The next step is the translation of scalar (J) coupling data into dihedral angle restraints and nuclear Overhauser enhancement data (NOEs) into interproton distance restraints. The final step is the calculation of a three-dimensional structure from these dihedral angle and interproton distance restraints by either distance geometry or restrained molecular dynamics. The strategy and the NMR experiments required to sequentially assign resonances differ depending on whether an unlabelled protein, a ^{15}N-labeled protein, a ^{13}C-labeled or a ^{15}N/^{13}C-labeled protein is available. Since the two IGF-I analogues in this study were ^{15}N-labeled proteins, the subsequent sections on the sequential assignment procedure and the structural restraints evaluation used for the three-dimensional structure calculations are specific for a ^{15}N-labeled protein.

The combination of two-dimensional homonuclear, two-dimensional heteronuclear and three-dimensional heteronuclear NMR experiments is utilised for obtaining sequence

specific assignments of a ^{15}N-labeled protein. This approach partially resolves the chemical shift degeneracy in two-dimensional homonuclear spectra by dispersing the resonances along an additional and better resolved heteronuclear frequency axis. These experiments that contain INEPT modules (Morris and Freeman, 1979) transfer magnetisation efficiently between the amide protons and a single ^{15}N nucleus via nearly conformationally independent one bond coupling constants. The relationship between 2D and 3D spectra is outlined in Figure 1.3. Each 3D plane at the ^{15}N shift of the amide nitrogen contains a subset of the 2D proton crosspeaks that are related to that particular amide group either by a dipolar or scalar interaction depending on the pulse sequence of the 3D experiment. The ^{15}N labels are incorporated into the protein by modern biotechnology methods and replace the naturally abundant ^{14}N nucleus which has less desirable NMR properties. The power of these experiments is manifested in the ability not only to increase resolution but to reveal connectivity paths between the protons and the nitrogen which can be manipulated by the various forms of the heteronuclear NMR experiments. For smaller proteins it is relatively rare to find that both the ^{1}H and ^{15}N chemical shifts of two H^N groups are degenerate and hence these heteronuclear experiments can be used in combination with the conventional homonuclear ones to resolve ambiguities. The 3D spectrum is viewed as a series of ^{1}H-H^N slices edited by the chemical shift of the directly bonded ^{15}N atom. Each ^{1}H-^{1}H plane represents the identical region covered by the 2D experiment, only containing fewer cross peaks. Connections between one residue and the next must be made not only between different sets of peaks but also between different planes of the spectrum (Marion et al., 1989).

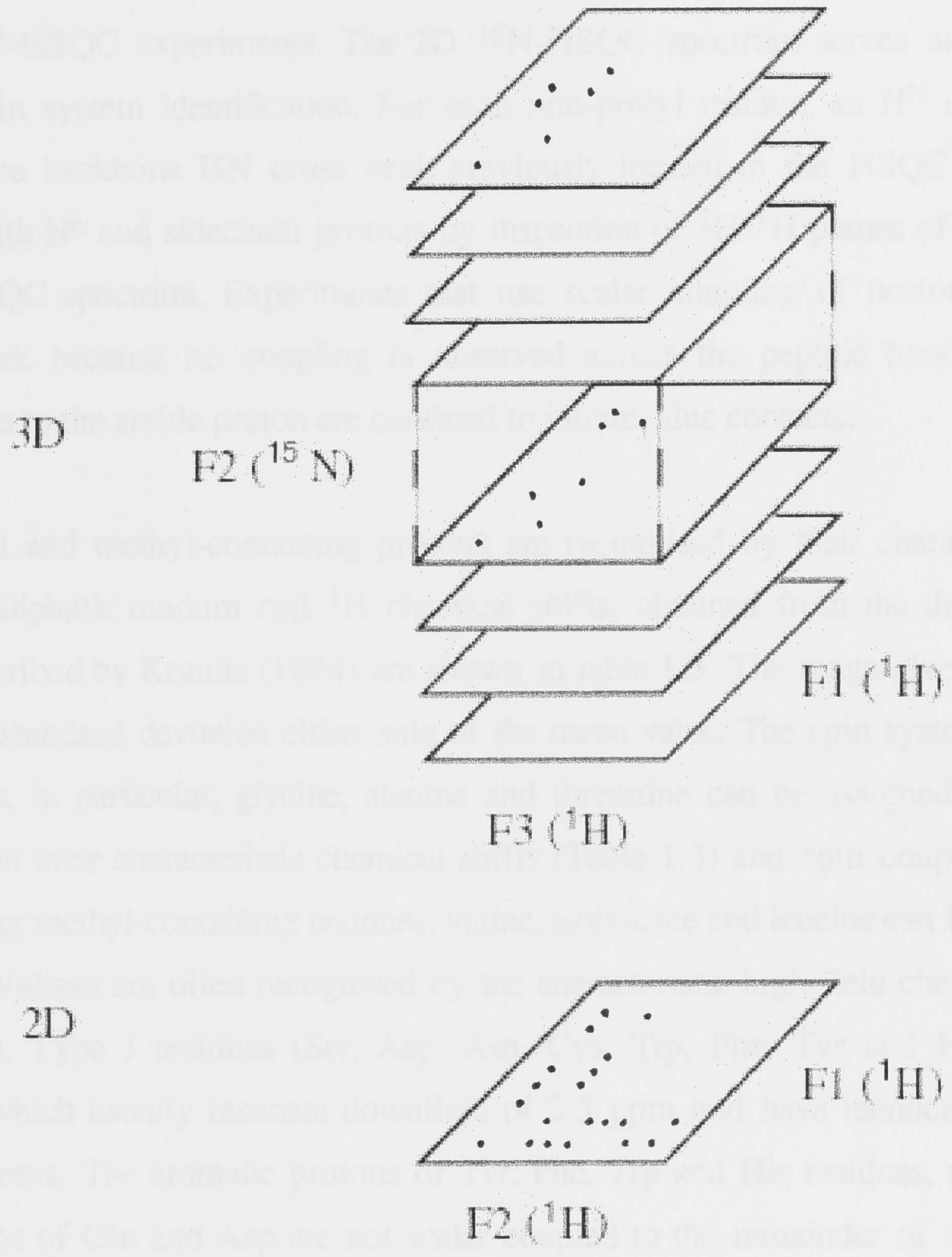

Figure 1.3: Schematic representation of the relationship between 3D heteronuclear NMR spectra and homonuclear 2D spectra. The 3D spectrum is viewed as a series of F1 (^{1}H) - F3 (H^{N}) slices edited by the chemical shift of the directly bonded ^{15}N atom along F2.

1.9.2 Sequential Assignments

The sequential assignment method (Wüthrich et al., 1982) consists of three stages. The first stage establishes the identity of the systems of spin-spin coupled resonances which belong to a particular amino acid residue type. The spin systems can be divided between short sidechain (Gly, Ala, and Thr), long sidechain (Val, Ile and Leu), Type J (Ser, Asp, Asn, Cys, Trp, Phe, Tyr and His) and Type U (Lys, Arg, Met, Gln, Glu and Pro). The spin systems are assigned an amino acid type based on coupling patterns and resonance chemical shifts. Assignment of an amino acid type is achieved by inspection of scalar coupling connectivities within individual residues using 2D ^{15}N-HSQC-TOCSY and 3D

^{15}N-TOCSY-HSQC experiments. The 2D ^{15}N-HSQC spectrum serves as the starting point for spin system identification. For each non-prolyl residue, an H^N cross peak is observed. The backbone HN cross peak previously located in the HSQC spectrum is correlated with H^α and sidechain protons by inspection of ^{1}H-^{1}H planes of the 3D ^{15}N-TOCSY-HSQC spectrum. Experiments that use scalar coupling of protons only, e.g., TOCSY, work because no coupling is observed across the peptide bond. Hence the connectivities to the amide proton are confined to intraresidue contacts.

Amide, alpha and methyl-containing protons are recognised by their characteristic shift values. The aliphatic random coil ^{1}H chemical shifts, obtained from the database of 13 proteins described by Kraulis (1994) are shown in table 1.3. The range given in the table extends one standard deviation either side of the mean value. The spin systems for some residue types, in particular, glycine, alanine and threonine can be assigned with relative ease based on their characteristic chemical shifts (Table 1.3) and spin coupling patterns. The remaining methyl-containing residues, valine, isoleucine and leucine can be difficult to distinguish. Valines are often recognised by the characteristic high field chemical shift of the α-proton. Type J residues (Ser, Asp, Asn, Cys, Trp, Phe, Tyr and His) have H^β resonances which usually resonate downfield of 2.5 ppm and have identical H^N-H^α-H^β spin subsystems. The aromatic protons of Tyr, Phe, Trp and His residues, and sidechain amide protons of Glu and Asp are not scalar coupled to the remainder of the sidechain. Association of the sidechain and backbone resonances of these spin systems has to be made on the basis of intraresidue NOE correlations. Type U amino acid residues (Lys, Arg, Met, Gln, Glu and Pro) all have two protons at the γ position which are coupled to the H^β protons. There may be some uncertainties among β, γ, and δ protons within one spin system of some of the Leu, Arg, Lys, Glu and Gln residues. These ambiguities can largely be resolved by obtaining 3D ^{15}N-TOCSY-HSQC spectra at two different mixing times. Overall, as a result of this first stage of the assignment procedure, the spin systems have either been assigned to a unique amino acid or spin system type. The subsequent steps assign the ^{1}H spin system residues to a unique amino acid in the sequence. The assignment of proline residues must be obtained from the 2D TOCSY and 2D NOESY, since the spin systems are not observed in the ^{15}N correlated spectra, as proline residues do not possess a free amide proton.

Table 1.3. Random coil ^{1}H chemical shift ranges (ppm) for the 20 common amino acids

Residue	αH	βH	Other
Gly	3.5-4.4		
Ala	3.8-4.8	1.1-1.6	
Val	3.6-4.8	1.6-2.2	γ 0.6-1.1
Ile	3.7-4.8	1.5-2.1	γ^1 1.0-1.7 γ^2 0.6-1.1 δ^1 0.5-1.0
Leu	3.9-4.9	1.4-2.0	γ 1.3-1.9 δ 0.5-1.0
Pro	4.0-4.9	1.4-2.5	γ 1.5-2.2 δ 3.2-4.1
Ser	4.2-5.1	3.6-4.2	
Thr	3.9-5.0	3.8-4.5	γ 0.9-1.4
Asp	4.1-5.0	2.4-3.0	
Glu	3.9-4.9	1.8-2.3	γ 2.1-2.5
Lys	3.8-4.7	1.4-2.0	γ 1.1-1.6 δ 1.4-1.9 ε 2.8-3.1
Arg	3.7-4.7	1.5-2.1	γ 1.3-1.8 δ 2.9-3.3
Asn	4.2-5.0	2.4-3.2	
Gln	3.8-4.8	1.6-2.3	γ 2.0-2.6
Met	4.0-5.0	1.8-2.2	γ 2.3-2.7 ε 1.6-2.7
Cys	3.8-5.3	2.6-3.3	
Trp	4.2-5.1	2.2-3.6	
Phe	4.1-5.2	2.6-3.4	
Tyr	4.2-5.2	2.5-3.3	
His	4.3-5.1	2.6-3.5	

The second stage of the assignment process is to cross the peptide bond and linearly connect as many spin systems as possible. In a ^{15}N-labeled protein there is no resolvable spin-spin coupling between protons of adjacent residues and therefore, through-bond correlation spectra cannot be used to determine the sequential connectivities. Sequential assignment is achieved by using through-space mediated (NOE) interactions to sequentially connect the spin systems identified from scalar correlations using 2D NOESY, 2D ^{15}N-HSQC-NOESY and 3D ^{15}N-NOESY-HSQC experiments. The primary assumption of the sequential assignment is that the probability is low that the amide proton of one residue is simultaneously close to the amide, α and β protons of any residue other than the preceding one. Individual long range NOEs to one of these protons may be

observed, but rarely to all three. In practice, sequential and intraresidue NOEs are usually more intense than long-range NOEs, which tend to be mostly weak. Identification of intense NOEs from the H^N of one spin system (i) to the H^N, H^α or H^β of a second preceding spin system (i-1) suggests that the two spin systems are adjacent in the primary sequence with the first spin system nearer to the C-terminus of the protein. Identification of a series of sequential NOE interactions places several spin systems in the order i, i +1, i +2, i +3, ...i + n.

The third stage is alignment in the sequence. As more spin systems are connected, the sequence of spin systems will eventually match a unique section of the primary amino acid sequence of the protein (which must be known prior to commencement of the assignment procedure). If the spin system types are well characterized (i.e. the majority of sidechain resonance positions have been identified), then alignment of four to five spin systems is usually sufficient to achieve sequence-specific assignment for those residues. A schematic diagram (Figure 1.4) of a dipeptide shows the possible short range NOEs used in the sequential assignment procedure.

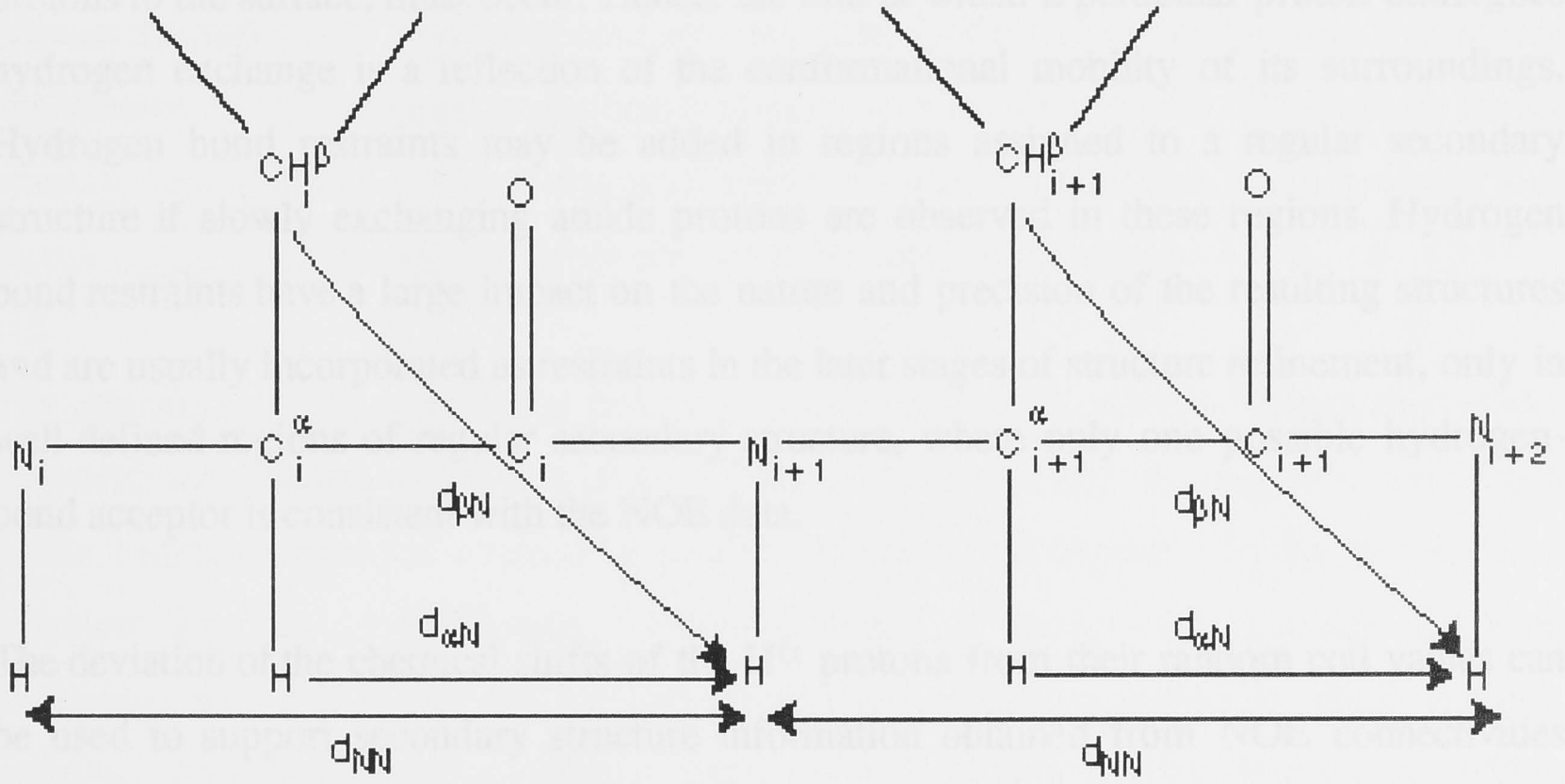

Figure 1.4: Dipeptide unit showing the sequential NOE connectivities, d_{NN}, $d_{\alpha N}$ and $d_{\beta N}$ used for sequential proton assignments.

1.9.3 Structure Determination from NMR Data

Secondary Structure

The secondary structure of a protein refers to the local conformation of its backbone. Regular polypeptide backbone folding patterns include helices, β-sheets and turns. Secondary structures in proteins have characteristic NOE patterns and scalar coupling constants (described in detail in section 4.1). Knowledge of the secondary structure elements in a protein assists with the structural determination process.

Hydrogen bonding has a major influence on the structure of proteins. The internal hydrogen bonds of a protein provide a structural basis for its native folding pattern. Weakly acidic protons, such as those of amide and hydroxyl groups (X-H) exchange with those of water. Protein interiors are largely excluded from contact with their surrounding aqueous solvent, and protons cannot exchange with solvent while they are engaged in hydrogen bonding. However, internal protons do exchange with solvent, indicating transient local unfolding that physically and chemically exposes these exchangeable protons to the surface, must occur. Hence, the rate at which a particular proton undergoes hydrogen exchange is a reflection of the conformational mobility of its surroundings. Hydrogen bond restraints may be added in regions assigned to a regular secondary structure if slowly exchanging amide protons are observed in these regions. Hydrogen bond restraints have a large impact on the nature and precision of the resulting structures and are usually incorporated as restraints in the later stages of structure refinement, only in well-defined regions of regular secondary structure, where only one possible hydrogen-bond acceptor is consistent with the NOE data.

The deviation of the chemical shifts of the H^{α} protons from their random coil values can be used to support secondary structure information obtained from NOE connectivities since H^{α} chemical shifts are dependent on the nature of secondary structure (discussed in detail in section 4.1). Thus, the H^{α} proton experiences an upfield shift with respect to the random coil value when in an α-helical conformation and a downfield shift when in a β-sheet conformation.

NMR Derived Structural Restraints

Chemical shifts, scalar coupling interactions and dipolar relaxation are sensitive to molecular conformation. Quantitative estimates of each of these parameters are used for structural analysis. The principal structural data from NMR used in determining protein structure are NOE cross peak intensities. The standard experiment for detecting through-space dipolar interactions is the 2D NOESY experiment, a 2D version of the NOE experiment (Overhauser, 1955), in which cross-peaks identify resonances from atoms near in space. The dipolar cross-relaxation rate is proportional to the inverse sixth power of the distance between two interacting protons and hence is sensitive only over short distances. In practice, the maximum interproton distance that can give rise to an observable first order NOE is 5Å. The sixth order dependence also permits the discrimination of proton pairs on the basis of internuclear distance. NOE cross peaks intensities are converted to interproton distances by direct calibration of the intensity of peaks observed in 3D ^{15}N-NOESY-HSQC spectra acquired at relatively short mixing times. The cross peak volume limits and upper bound distances for the categories are estimated from the NOE intensities observed for protons of known covalent geometry (e.g. geminal methylene protons) or between protons in regions of regular secondary structure. Conservative upper-bound distances compensate for cross-peak intensities that may be affected by spin diffusion or partial overlap, both of which can lead to unjustifiably tight distance restraints. Intensities can also be calculated and converted to upper bounds in an automated fashion.

The magnitude of $^{3}J_{HNH\alpha}$ scalar coupling constants contain valuable information about the backbone dihedral angles in a protein. The magnitude of the $^{3}J_{HNH\alpha}$ values is related to the backbone dihedral angle of a protein, θ (shown in Figure 1.5) by the Karplus equation (Karplus, 1959; Bystrov, 1976) (discussed in section 5.3.2) which has been parameterised for this purpose by correlating the J values measured in small globular proteins with dihedral angles obtained from crystallographic studies (Pardi et al., 1984; Ludvigsen et al., 1991). Dihedral angle restraints are used in calculating the protein structure. $^{3}J_{HNH\alpha}$ values can be used to directly determine secondary structure whereas $J_{N\beta}$ and $J_{\alpha\beta}$ values are used to identify prochiral groups, i.e. sidechain dihedral angles. The latter are particularly important for buried sidechains that take up fixed positions. Surface sidechains frequently are mobile and averaged $J_{N\beta}$ and $J_{\alpha\beta}$ values are indicative of sidechain conformational averaging.

Calculating Protein Structures from NMR Data

A variety of methods are available for calculating protein structures using NMR-derived restraints. These methods determine coordinates for the atoms in the protein that will satisfy the input distance and dihedral angle restraints while exploring all regions of conformational space compatible with the NMR parameters. Since the restraints are included as ranges of allowed values, the data contain experimental uncertainties and because not all possible distance, dihedral and hydrogen bonding restraints are observable, the structure calculation is repeated to determine an ensemble of low-energy structures consistent with the input data. The low-energy ensemble of structures should minimise violations of the input restraints and minimise the root-mean-square deviation (RMSD) between members of the ensemble.

The two most common approaches for generating structures are distance geometry (DG) and restrained molecular dynamics (rMD), or a method employing both. Distance geometry determines ensembles of 3D structures consistent with an incomplete set of distance constraints. Not all distances between all the possible proton pairs can be used because the NOE effect is limited to distances less than 5Å and even these distance restraints are not known precisely. In the rMD technique, molecular dynamics force fields are supplemented by pseudo-energy terms based on the NMR-derived restraints. The potentials drive the structure toward a conformation that will reduce the violation of the restraints during a forced heating followed by a cooling annealing cycle. The most efficient implementations of the rMD method use a simplified force field in which the bond length, the bond angle and the repulsive van der Waals terms are retained. These implementations are referred to as dynamic simulated annealing. For this thesis, structures were generated using the computer program, DYANA 1.5 which employs simulated annealing combined with molecular dynamics in torsion angle space. An iterative process of calculation is employed in which successive rounds of structure calculations are used to make further NOE assignments.

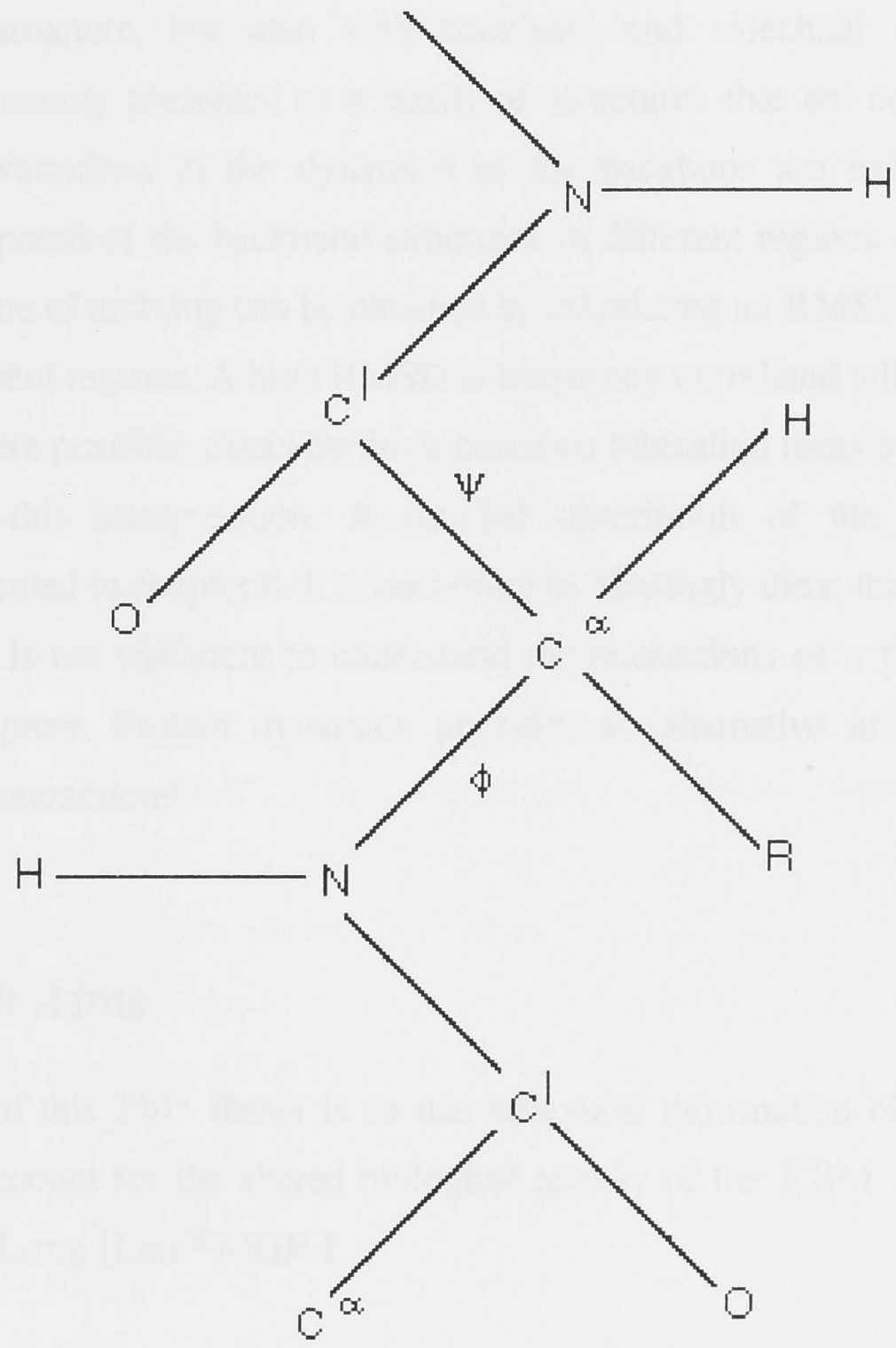

Figure 1.5: Protein backbone with the backbone dihedral angles ϕ and ψ indicated. The amino acid sidechain is represented by R.

The precision to which a structure can be calculated is directly proportional to the number of experimental restraints used to generate it. The use of ^{15}N-labelling allows many more ^{1}H-^{1}H NOEs to be unambiguously identified. In addition, extra data are available to stereospecifically assign prochiral groups, for example, $C^{\beta}H_2$ methylene protons, γ-CH_3 and δ-CH_3 groups of valine and leucine residues, respectively, which helps in obtaining structures of high precision.

NMR Structures

Because NMR structures are obtained in the solution state, there is not only a direct correlation with structure, but also with backbone and sidechain dynamics. NMR structures are commonly presented as a family of structures that are consistent with the input restraints. Variations in the dynamics of the backbone are indirectly related to variations in the spread of the backbone structures in different regions of the protein. A quantitative estimate of mobility can be obtained by calculating an RMSD of the family of structures of different regions. A high RMSD is frequently correlated with regions of high flexibility, but where possible, direct methods based on relaxation measurements should be used to confirm this interpretation. A detailed description of the direct relaxation techniques is presented in chapter 6. It is becoming increasingly clear that knowledge of a protein's structure is not sufficient to understand the interactions of a protein with other proteins and receptors. Protein dynamics provides an alternative and complementary insight into these interactions.

1.10 Research Aims

The primary aim of this PhD thesis is to use structural information obtained by NMR spectroscopy to account for the altered biological activity of the IGF-I analogues, Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I.

As discussed in section 1.8, Long-[Arg^3]-IGF-I is one of the most potent of the recombinant fusion protein analogues of IGF-I in which the N-terminal region has been modified (Francis et al., 1992). It has been postulated that the increased potency of Long-[Arg^3]-IGF-I is due to reduced binding to IGFBPs rather than increased affinity for the cell receptors. It is unclear whether the increased potency of Long-[Arg^3]-IGF-I is a direct result of the N-terminal extension hindering binding to IGFBPs or because the N-terminal extension disrupts the secondary and tertiary structure of regions of the IGF-I domain involved in IGFBP binding.

The three-dimensional structure elucidation of Long-[Arg^3]-IGF-I will determine whether the secondary structural elements of IGF-I, namely the three helices and their relative orientation, have been retained in Long-[Arg^3]-IGF-I. Chemical shift index analysis of H^{α}

will provide information as to whether any residues of Long-[Arg^3]-IGF-I exist in different chemical environments when compared with IGF-I. Chemical shift index, $^3J_{NH\alpha}$ values and NOE patterns will be used to identify regions of well defined secondary structure in Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I. Amide proton exchange rates will indicate which residues are at the surface of the protein and which are protected from exchange. This information will be compared with HN exchange data for IGF-I and will provide insight into whether the sites of IGF-I involved in IGFBP or receptor binding remain solvent accessible in Long-[Arg^3]-IGF-I. The three dimensional structure of Long-[Arg^3]-IGF-I will be determined with a particular focus on the N-terminal extension and how it interacts with the IGF-I domain of Long-[Arg^3]-IGF-I. Backbone dynamics studies of Long-[Arg^3]-IGF-I will also be performed to gain insight into the degree of mobility of regions of Long-[Arg^3]-IGF-I critical for IGFBP or receptor binding. This information will be used to rationalise the increased biological potency of Long-[Arg^3]-IGF-I.

A similar approach will be taken in examining Long-[Leu^{60}]-IGF-I. However, the global fold of Long-[Leu^{60}]-IGF-I will not be calculated unless a significantly different secondary structure is detected compared to Long-[Arg^3]-IGF-I. Structural parameters of Long-[Leu^{60}]-IGF-I will be compared with IGF-I and Long-[Arg^3]-IGF-I and interpreted in terms of binding affinities for IGFBPs and receptors and biological potency. In order to accurately elucidate the residues involved in IGFBP binding, chemical shift mapping studies of Long-[Leu^{60}]-IGF-I with IGFBP-2 will be performed. This involves acquiring a series of 2D ^{15}N-HSQC spectra without IGFBP-2 and subsequently with increasing concentrations of IGFBP-2, enabling observations of perturbed resonances which are implicated in the binding interaction.

When combined together, regions of secondary structure, amide proton exchange rates, 3D structural data, backbone dynamics and chemical shift mapping will provide insight into the regions of IGF-I and the IGF-I analogues important for IGFBP and receptor binding. These results will be interpreted in terms of enhanced biological activity of the analogues and will, in the longer term, assist in the design of additional potent IGF-I analogues.

CHAPTER 2

Protein Expression and Purification

2.1 Introduction

The two IGF-I analogues to be characterized by NMR are approximately 9 kDa in mass, which is at the limit of structure determination by homonuclear NMR methods alone (Oschkinat et al., 1994). This limitation is particularly acute for proteins with a high helical content, in which the majority of the alpha protons are expected to resonate within a range of just one part per million upfield of the water resonance in a crowded region of the ^{1}H NMR spectrum. The three-dimensional structure of IGF-I was determined by two independent groups using purely homonuclear techniques (Cooke et al., 1991; Sato et al., 1993). The analogues studied in this thesis, Long-[Arg^{3}]-IGF-I and Long-[Leu^{60}]-IGF-I are larger than IGF-I itself, each with an additional 13 amino acids, resulting in further overlap in the ^{1}H NMR spectrum in comparison with IGF-I.

Some of the ^{1}H resonance overlap can be alleviated in ^{15}N-labeled proteins by ^{15}N-separated NMR spectroscopy. Acquisition of ^{15}N-separated multi-dimensional spectra is only feasible at normal protein concentrations with a ^{15}N-labeled protein. ^{15}N-separated spectra segregate ^{1}H resonances into a series of planes in a third dimension according to the ^{15}N resonance frequency of the directly bonded amide nitrogen. The efficacy of using the ^{15}N-labeled protein has been improved by the reduction in price of ^{15}N source ($^{15}NH_4Cl$). Both proteins were expressed recombinantly in *E. coli* on minimal media with $^{15}NH_4Cl$ as the sole nitrogen source employing the methods developed by King et al. (1992).

Uniformly ^{15}N-labeled Long-[Arg^{3}]-IGF-I had been expressed and purified by Gary Shooter (Department of Biochemistry, University of Adelaide) prior to the commencement of my PhD project, following the methods outlined by King et al. (1992) and as such, these methods and results will not be discussed in this chapter. The expression and purification of Long-[Leu^{60}]-IGF-I, described in the subsequent sections of this chapter, also followed the general methods of King et al. (1992) and was performed initially under the guidance of Dr Steve Milner at the CRC for Tissue Growth and Repair in Adelaide and continued unsupervised at the Research School of Chemistry, ANU.

2.2 Materials

2.2.1 Expression Vector Construct

E. coli strain JM101 was transformed with the p[Met^{1}]-pGH(11)-Val-Asn-[Leu^{60}]-IGF-I expression vector and maintained on minimal medium as described by King et al. (1992).

2.2.2 Chemicals

($^{15}NH_4$)Cl (99% enrichment) was obtained from Cambridge Isotope Laboratories (Cambridge, MA). The media components included $^{15}NH_4Cl$, K_2SO_4, KH_2PO_4, Na_2PO_4, glucose and $MgSO_4$, thiamine and ampicillin. Isopropyl β-D-thiogalactopyranoside (IPTG) was used for induction. Trace elements included $FeSO_4$, $MnSO_4$, $ZnSO_4$, $CuSO_4$, sodium citrate, and concentrated HCl. For purification of Long-[Leu^{60}]-IGF-I, urea was obtained from Merck, Tris, 2-hydroxyethyldisulphide and LiOH from Aldrich, EDTA and acetic acid from Ajax, $ZnCl_2$ from Sigma, DTT from Boehringer Mannheim and glycine, ammonium acetate, NaCl, TFA, acetonitrile, NaOH, and HCl from BDH.

2.2.3 Separation and Analytical Columns

Initial purification of the IGF analogues was by ion exchange chromatography using a Pharmacia column of dimensions 400 mm x 30 mm packed with SP Sepharose Fast Flow (FFS) (Pharmacia-Biotech) media, a strong cation exchange gel. The pump was a Delta Prep 4000 (Waters) with a Millipore Lamda Max absorbance detector (Waters) measuring absorbance at 280 nm connected to a Linear Instruments model 555 chart recorder. Subsequent purification was by reverse phase HPLC. A Waters C_4 column was used with a pore size of 300 Å and dimensions of 40 mm x 100 mm. The pump and system controller used were a Waters Delta Prep 4000 and Waters 4000 System controller with a Waters PrepPack RCM Base. Detection was with a Waters Millipore Lamda Max absorbance detector monitored with Maxima 820 software. The purity of the column fractions was monitored by analytical HPLC using a SGE C_4 column of dimensions 4 mm x 250 mm, pore size 300 Å, 15 µm particle size and nucleosil packing. A Waters 510 HPLC pump with a Rheodyne Injection port was used with a Waters Millipore Lamda Max absorbance detector and the system controller was Maxima 820 software.

2.2.4 Spectroscopy and Sequencing

The concentration of the NMR sample was determined by UV-visible spectrophotometry using a Cary 1E UV-Visible Spectrophotometer with Cary E Software on an IBM PC central controller. Mass spectrometric analysis was performed on a Fisons Instruments QUATTRO II Triple Quadrupole Mass Spectrometer. The protein was submitted to the Biomolecular Resource Facility at the John Curtin School of Medical Research where N-terminal sequencing of the protein was performed on a Model 494 PROCISE-HT protein sequencer in pulsed-liquid mode (Perkin Elmer Applied Biosystems, Foster City, CA).

2.3 Methods

2.3.1 Overexpression of Long-[Leu^{60}]-IGF-I

Preparation of Inoculum

The *E. coli* culture containing the Long-[Leu^{60}]-IGF-I construct was plated out on minimal media two days prior to the planned commencement of fermentation. A shaker culture was prepared the same morning that fermentation commencement was planned. Ampicillin (200 µg/ml) was added to 200 ml of Luria Broth (Ausubel et al., 1998) in the shaker flask. To this, a single colony of organisms was added to the culture which was shaken at room temperature. The optical density of the culture was monitored throughout the day.

Fermentation

The base media for the fermentations was prepared with the following salts: 7 mM K_2SO_4, 12 mM KH_2PO_4, and 18 mM Na_2HPO_4. The salts were dissolved in 1.6 L of water filtered through a 1 µm filter and the pH adjusted to 7.0. D-glucose (139 mM) and $MgSO_4$ (2.3 mM) were dissolved in 0.4 L of water and filtered through a 1 µm filter. Both solutions were autoclaved along with the fermentation assembly (Applikon 5 L fermenter). 4 ml of trace element stock solution was sterile filtered through a 0.22 µm filter and added to the fermenter. The trace element stock solution contained 5 g/L Fe_2SO_4, 1.28 g/L $MnSO_4$, 2.16 g/L $ZnSO_4$, 0.2 g/L $CuSO_4$, 22 g/L sodium citrate and 10 g/L concentrated HCl. A stock solution of thiamine (0.3 g in 7.5 ml) was made and sterile-

filtered through a 0.22 µm filter and 2 ml was added to the fermenter. 1 ml of a 100 mg/ml ampicillin solution was also sterile filtered and added to the fermenter. 3.21 g $^{15}NH_4Cl$ (30 mM) was dissolved in 10 ml water and sterile-filtered. This solution was added to the fermenter through the septum immediately prior to inoculation.

The shaker culture was left during the day at room temperature and inoculated when the organism populations had reached the doubling phase. The inoculum volume was 900 µl. The optical density of the fermentation was measured throughout the day, the temperature was maintained at 37 °C, dissolved oxygen (initially set to 60%) was monitored and pH was maintained at 7.0. The stirrer motors were set at 600 rpm. Once the dissolved O_2 dropped below 60%, the air flow was increased. The fermenter control software, FC4, was manufactured by Real Time Engineering, NSW. The fermentation was induced with IPTG (0.67 ml of a 1 M solution) when the optical density was 1.5. The fermentation was terminated when the optical density reached a reading that remained stable, indicating that exponential growth was no longer occurring.

Homogenisation

Following the termination of the fermentation, the broth was transferred to a pre-cooled stainless steel holding tank using a masterflex peristaltic pump. The broth was homogenised by four passes at 10000 psi using a Rannie bench top homogeniser model 8.30H.

Centrifugation

Following homogenisation, the cells were harvested by centrifugation using a Beckman J2-21 centrifuge. The homogenate was distributed into 6 x 500 ml Beckman centrifuge bottles which were balanced so that the pairs were within 1 g of each other. Initially, the homogenate was centrifuged at 10000 rpm at 4 °C for 25 min. This spin was performed to reduce the effective volume, retain all the inclusion bodies, but this also included cell debri. The supernatant was decanted off into a plastic bottle and retained until the purification was complete.

The inclusion body pellets in each of the centrifuge bottles were resuspended in 200 ml of inclusion bodies' washing buffer (30 mM NaCl, 10 mM KH_2PO_4, 0.5 mM $ZnCl_2$) and

centrifuged at 6500 rpm at 4 °C for 25 min. This spin was designed to retain the largest inclusion bodies but leave most of the cell debri in suspension. The supernatant was decanted off and again retained.

The inclusion body pellets were again resuspended in 200 ml of inclusion bodies' washing buffer and centrifuged at 10000 rpm at 4 °C for 25 min. This spin was designed to obtain a solid inclusion bodies pellet. The supernatant was decanted off and retained.

The inclusion body paste was scraped from each of the centrifuge bottles and the yield of wet inclusion bodies recorded. Different temperature compensation settings were required for different rotor speeds. The supernatant retained from the centrifugation was tested by HPLC to ensure that there was no protein present before discarding.

2.3.2 Purification of Long-[Leu^{60}]-IGF-I

Dissolution of Inclusion Bodies

Inclusion bodies were resuspended at 10% (w/v) in a solution of 8 M deionized urea, 50 mM glycine, 0.5 mM $ZnCl_2$, 16 mM DTT (dithiothreitol) and 0.1% (w/v) Tris. The dissolution solution was adjusted to pH 9.0 with concentrated HCl. The resuspended inclusion bodies were left to stand at room temperature for one hour.

Refolding of Reduced Long-[Leu^{60}]-IGF-I

Refolding of Long-[Leu^{60}]-IGF-I was started immediately following dissolution. The refolding mixture contained 4 M urea, 0.1 M Tris, 20 mM glycine, 0.5 mM $ZnCl_2$, 0.4 mM DTT and 1 mM 2-Hydroxyethyldisulphide at pH 9.0. The dissolved protein was 0.025% v/v of the total refolding volume. 2-hydroxyethyldisulphide was added quickly once the dissolved protein had been added to the remaining refold mixture at pH 9.0, with rapid stirring to avoid aggregation of the protein. The refolding reaction was left to proceed at room temperature for 2-3 hours with continual stirring and was monitored at least three times during the refold by analytical HPLC. The refolding reaction was terminated by adding 1.2 ml of concentrated HCl for each 100 ml of the refolding mixture.

Ion Exchange

The column used for Ion Exchange was sanitised and equilibrated by performing the following washes at 25 ml/min; 500 ml ethanol, 500 ml milli-Q water, 250 ml 0.5 M NaOH, 500 ml sterile water, 500 ml acetic acid, 500 ml Buffer A (8 M urea, 50 mM ammonium acetate, pH 4.8) and 500 ml Buffer B (8 M urea, 50 mM ammonium acetate, pH 4.8, 2M NaCl). The protein was loaded onto the column and washed free of UV absorbing material with several column volumes of Buffer A. Long-[Leu^{60}]-IGF-I was eluted using a single step gradient, 100% Buffer A to 100% Buffer B. The gradient was repeated to ensure all protein had been eluted from the column. The protein, which was detected by an increase in absorbance at 280 nm, was collected as a single fraction. The fraction containing the protein was analysed by analytical reverse-phase HPLC. All effluent was retained and analysed by analytical reverse-phase HPLC to verify that it contained no protein before further purification was performed on the collected fraction.

Reverse-Phase HPLC Purification of Long-[Leu^{60}]-IGF-I

The refolded, Long-[Leu^{60}]-IGF-I was purified by reverse phase HPLC chromatography on a C4 Prep-Pak column (40 mm diameter x 100 mm; 300 Å pore size, 15 µm particle size; Millipore-Waters) in a Waters PrepPak RCM Base at 7500 psi, using a Waters Delta Prep 4000 Preparative Chromatography System with a Waters 4000 System Controller. The column was sanitised with 600 ml 100% methanol, 500 ml 0.1% TFA/40% acetonitrile, 500 ml 0.08% TFA/80% acetonitrile, and 500 ml 0.1% TFA at a flow rate of 25 ml/min. A gradient (0-80% acetonitrile in 20 min) was run to remove any UV absorbing material, followed by 0.1% TFA for 10 min. The protein was then loaded onto the column at a flow rate of 25 ml/min. The protein was eluted with the following gradient; 0-20% acetonitrile in 10 min, then to 40% acetonitrile in 210 min and to 64% acetonitrile in 211 min) as monitored by absorbance at 280 nm. The fractions were pooled for analysis by HPLC.

In order to separate the desired protein from the smaller components, the protein was purified by reverse-phase HPLC on a C4 column employing an n-propanol gradient with heptaflourobutyric acid (HFBA) as the counter ion. The pooled fractions were loaded onto the column and eluted with a gradient, 20-35% n-propanol in 154 min at a flow rate of 10 ml/min.

2.3.3 HPLC Analysis

Analysis of the protein concentration and the purity of the sample were performed by analytical HPLC at each step of the purification process. Generally 4-8 µg of protein in 100 µl of liquid was injected onto the column. For the analysis of dissolved inclusion bodies, 20 µl was diluted to 1 ml with 0.1% TFA and 100 µl was injected onto the column. Analysis of the protein content in the refolding buffer was performed by diluting 150 µl of refolding mixture to 500 µl of 1.5% TFA and 250 µl was injected. Three samples were taken, at the commencement of refolding, after 90 minutes of refolding and at the termination of refolding after approximately two hours. Following ion exchange chromatography, the fractions containing protein were analysed by HPLC by diluting 20 µl to 1 ml with 0.1% TFA and injecting 250 µl onto the column. The effluent was checked prior to discarding by diluting 150 µl to 500 µl 1.5% TFA and injecting 250 µl onto the column. Following reverse phase HPLC of Long-[Leu^{60}]-IGF-I, 20 µl was taken from each fraction containing protein to give a total volume of 100 µl. This was diluted to 400 µl with 0.1% TFA, and 250 µl was injected onto the column.

For analytical HPLC analysis, Buffer A was 0.1% TFA and Buffer B was 80% CH_3CN and 0.08% TFA. The gradient run was 20% to 80% CH_3CN in 20 min for all runs which is 25% Buffer B to 100% Buffer B in 20 min; the eluent was held at 100% Buffer B until 30 min and then decreased to 25% Buffer B in 1 min.

2.3.4 Preparation of NMR Sample

The target sample concentration was 1 mM, which was achieved by weighing 4.6 mg of Long-[Leu^{60}]-IGF-I and dissolving in 500 µl 90% H_2O / 10% D_2O. The pH was measured to be 3.3. The protein concentration was verified by UV absorption as discussed in section 2.3.5.

2.3.5 Calculation of Protein Concentration

The amount and purity of the protein Long-[Leu^{60}]-IGF-I were determined by comparing the spectrophotometric absorbance at 214 nm with that of reference IGF-I by reverse-phase HPLC chromatography on a C4 microbore column as described in section 2.3.3. The following relationship was used to estimate the amount of protein, where Ext. Coeff. refers to the extinction coefficient.

(Area of Peak / Ext. Coeff.) x dilution factor x (1000 µl / injection volume µl) (2.1)

The absorbance at 280 nm was used to accurately determine the protein concentration of the NMR sample. From the NMR sample, 100 µl was taken and diluted to 1 ml and the absorbance at 280 nm measured. The molar extinction coefficient of Long-[Leu^{60}]-IGF-I at 280 nm is 3270 $M^{-1}cm^{-1}$ calculated using the method of Gill and von Hippel (1989). The concentration was determined from the relationship, $A = \varepsilon cl$, where A is the absorbance at 280 nm, ε is the extinction coefficient in $M^{-1}cm^{-1}$, c is the concentration (M) and l is the cell path length, 1 cm.

2.3.6 Mass Spectrometry

Positive electrospray ionisation was used to determine the mass of the protein fraction in order to verify the percentage incorporation of ^{15}N into Long-[Leu^{60}]-IGF-I. The protein sample concentration was 0.2 mM in H_2O and 20 µl was injected at a solvent flow rate of 10 µl/min. The raw data was transformed using the standard instrument software, MASSLYNX.

2.4 Results

2.4.1 Overexpression of Long-[Leu^{60}]-IGF-I

The growth of cells was induced when the optical density had reached 1.5. Exponential growth continued until the optical density reached 7. The fermentation was terminated after two measurements were taken with the same optical density value, indicating no further growth. The pH did not rise significantly, as occurs when glucose becomes exhausted and nitrogen is still available, which is the case for the production of unlabelled protein using this procedure. Because nitrogen was exhausted, the inclusion body wet mass yield (6.16g) was lower than usually obtained for the same protein with unlimited nitrogen present.

2.4.2 Purification of Long-[Leu^{60}]-IGF-I

The HPLC profile of the dissolved inclusion bodies is shown in Figure 2.1. Two major peaks are observed in the HPLC profile. The first peak, which elutes at approximately 18 min contains the desired protein. The second peak is most likely an N-formyl-methionine

derivative of the fusion protein, a result of incomplete processing by *E. coli* which has previously been observed (Milner et al., submitted for publication).

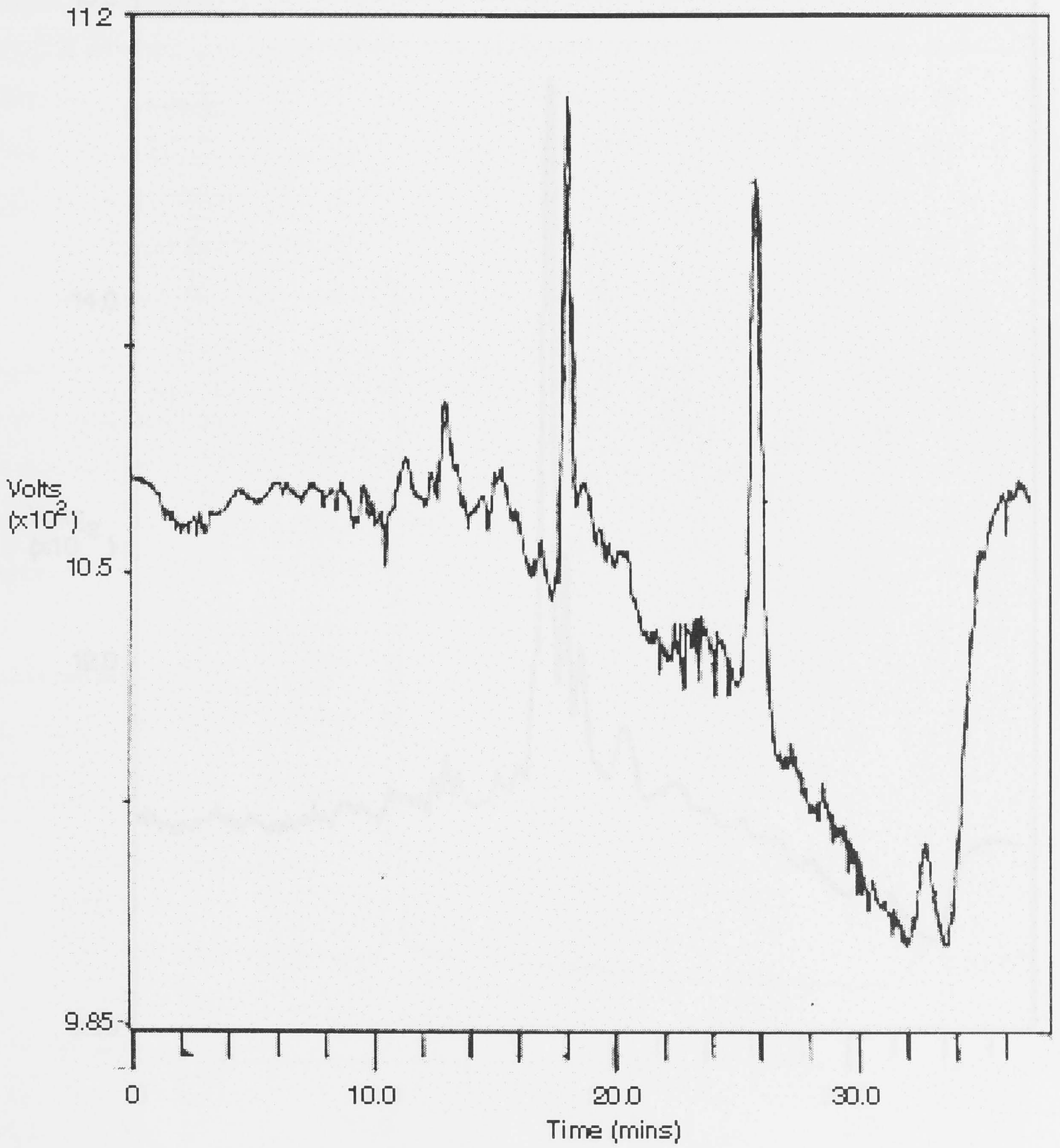

Figure 2.1: HPLC trace of Long-[Leu^{60}]-IGF-I inclusion bodies

The HPLC profile of the fractions containing the refolded protein are shown in Figure 2.2. Although the HPLC profile is not clean, there was one major peak present arising from the protein of interest which was collected. This peak was then purified with an ion exchange column. The protein eluted in one repeat of the gradient. Reverse phase HPLC purification of Long-[Leu^{60}]-IGF-I was then performed using the TFA/CH_3CN buffer system followed by purification using a HFBA/n-propanol buffer system. The HPLC trace of the

purified protein is shown in Figure 2.3 and it is apparent that the sample is clean as a result of these steps.

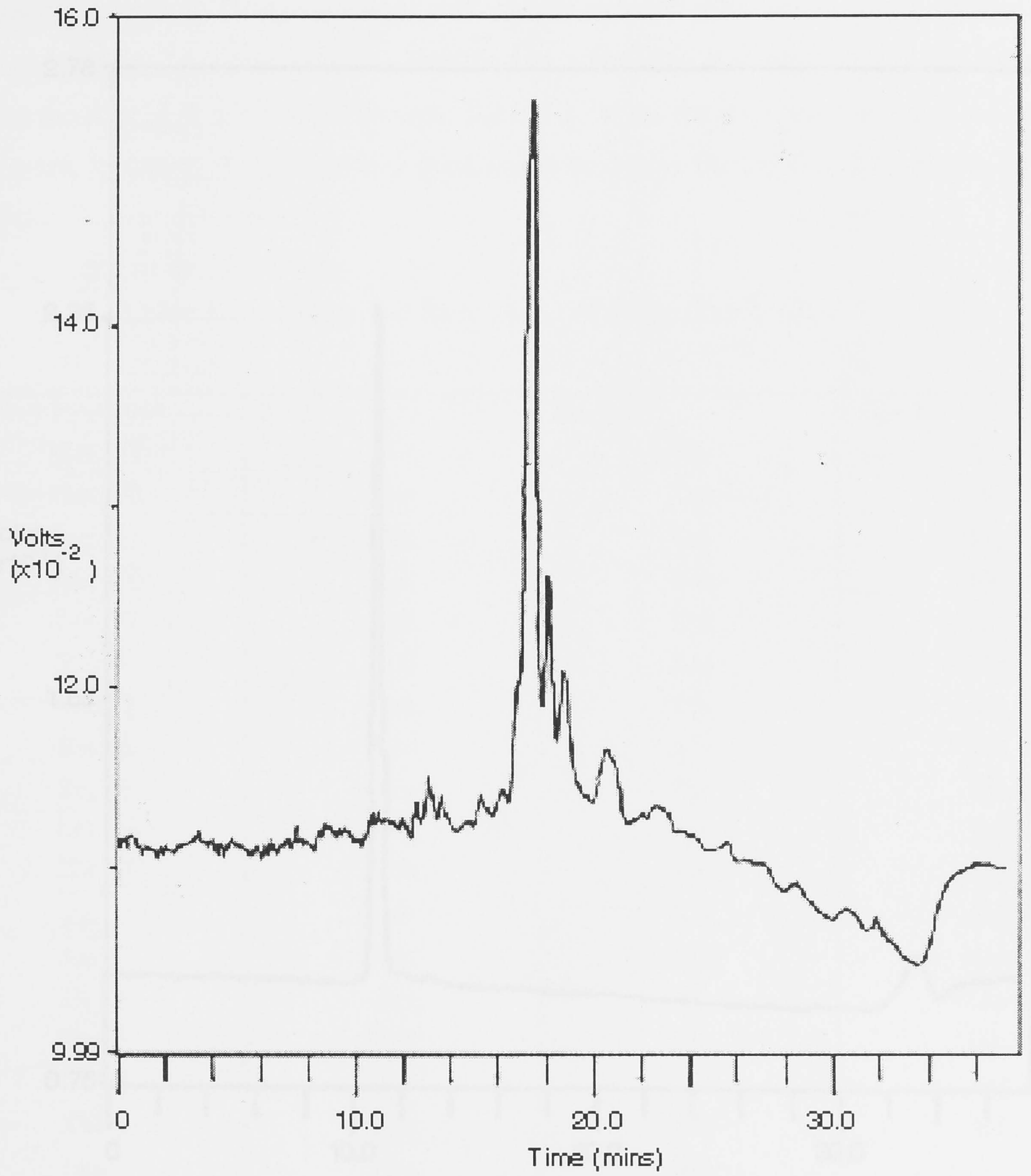

Figure 2.2: HPLC profile of refolded Long-[Leu^{60}]-IGF-I following ion exchange chromatography.

2.4.3 Yield and Sample Concentration

Prior to freeze drying the fractions containing protein, the amount of protein was estimated by analytical HPLC to be approximately 6 mg. The protein was freeze dried and an NMR sample made up to approximately 1 mM. The concentration was determined by measuring

the absorbance at 280 nm. The absorbance at 280 nm of a 1 in 10 dilution of the NMR sample at 280 nm was 0.4378 Absorbance Units. Using equation 2.1, the concentration of the NMR sample in 90% H_2O/10% D_2O was determined to be 0.96 mM.

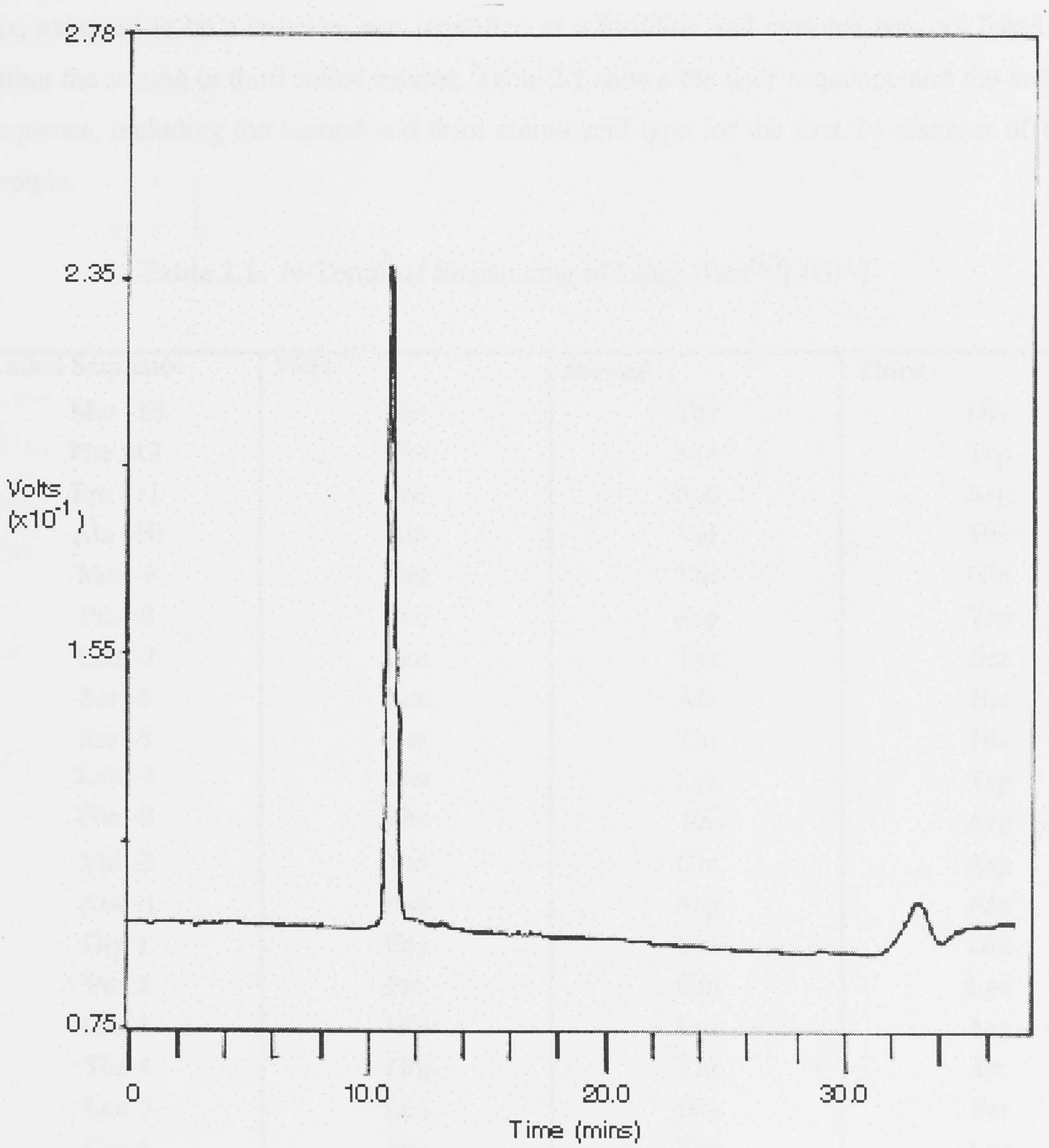

Figure 2.3: HPLC trace of purified Long-[Leu^{60}]-IGF-I

2.4.4 N-terminal Sequence Analysis

N-terminal sequence analysis was performed for the first 24 amino acid residues, i.e. the entire N-terminal extension and the first 11 residues of the IGF-I domain (Met^{-13} to Val^{11}). The determined sequence was MFPAMPLSSLFVNGPEELHGAALV, which compares favourably with the expected sequence MFPAMPLSSLFVNGPETLCGAELV.

Residues four, six and nine differed in composition between the two sequences. Residue four, expected to be threonine was identified as a glutamic acid, although threonine was the second called amino acid type. Similarly, residue nine, expected to be glutamic acid was identified as alanine, though glutamic acid was the second called amino acid type. Residue six, expected to be a cysteine, was identified as a histidine and cysteine was not listed as either the second or third called residue. Table 2.1 shows the user sequence and the called sequence, including the second and third amino acid type for the first 24 residues of the protein.

Table 2.1. N-Terminal Sequencing of Long-[Leu^{60}]-IGF-I

Called Sequence	First	Second	Third
Met -13	Met	Thr	Gly
Phe -12	Phe	Asp	Trp
Pro -11	Pro	Asp	Arg
Ala -10	Ala	Val	His
Met -9	Met	Thr	Gln
Pro -8	Pro	Asp	Trp
Leu -7	Leu	Tyr	Ser
Ser -6	Ser	Ala	His
Ser -5	Ser	Thr	His
Leu -4	Leu	Lys	Trp
Phe -3	Phe	Ile	Arg
Val -2	Val	Gln	Asp
Asn -1	Asn	Asp	Ala
Gly 1	Gly	Lys	Glu
Pro 2	Pro	Gln	Lys
Glu 3	Glu	Lys	Arg
Thr 4	Glu	Thr	Ile
Leu 5	Leu	Glu	Ser
Cys 6	His	Trp	Lys
Gly 7	Gly	Ala	Asp
Ala 8	Ala	Lys	Gly
Glu 9	Ala	Glu	Arg
Leu 10	Leu	Glu	Arg
Val 11	Val	Trp	Leu

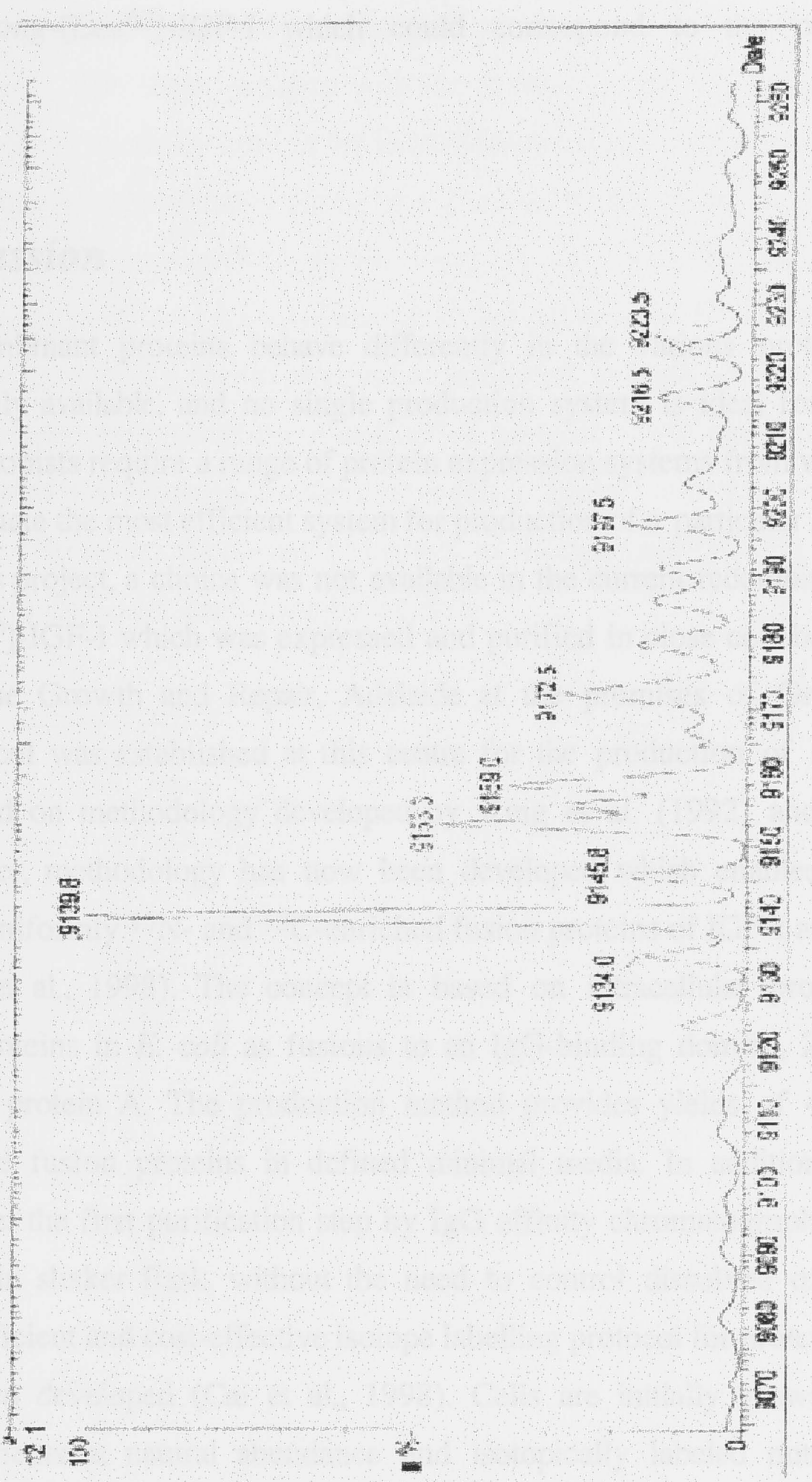

Figure 2.4: Mass Spectrum of Long-[Leu^{60}]-IGF-I

2.4.5 Mass Spectrometry

Mass spectrometric analysis of the pooled fractions showed a major species with a mass of 9138.23 ± 0.11 Da which is consistent with the expected mass of 100% ^{15}N-labeled Long-[Leu^{60}]-IGF-I. The mass spectrum (Figure 2.4) shows another smaller peak with a mass of 9154.47 ± 0.10 Da. This is an increase in molecular mass of 16 Da compared to

^{15}N-labeled Long-[Leu^{60}]-IGF-I which could correspond to the oxidation of a methionine.

2.5 *Discussion*

Different recombinant proteins behave differently in the various protein production systems currently available, and no single production system is ideal for every protein. NMR spectroscopists require a range of protein expression systems from which to choose in order to evaluate the most efficient system for production of a particular protein. Due to the scope of this project, a choice was not available in the current recombinant production of Long-[Leu^{60}]-IGF-I which was expressed and purified in close collaboration with the CRC for Tissue Growth and Repair, Adelaide at the premises of GroPep Pty Ltd. Standard protocol was established at this centre for the production of IGFs and their analogues based on methodology developed by King et al. (1992) and Francis et al. (1992). However, methodology has now been developed which enables the high-level production of uniformly ^{15}N- and ^{13}C-enriched fusion proteins of IGF-I and IGF-II in *E. coli* (Jansson et al., 1998). The concept is based on intracellular production of the recombinant proteins in *E. coli* as fusions to an IgG-binding domain, Z, derived from Staphylococcal protein A. The production method provides yields of 40-200 mg/l of isotope-enriched fusion proteins in defined minimal media. In addition, the Z fusion partner facilitates the first purification step by IgG affinity chromatography. These yields were obtained in shaker flask, without the need to control dissolved oxygen and pH. Recently, an efficient and cost-effective isotope labelling protocol for proteins expressed in *E. coli* has been developed (Cai et al., 1998). Cells are initially grown in a medium containing nutrients at natural abundance and isotopically labeled nutrients are only supplied at the later stages of growth and during protein expression. This permits the accumulation of a large cell mass without the need to employ large amounts of expensive isotopically labeled nutrients. This method was not available at the commencement of the production of Long-[Leu^{60}]-IGF-I. The aim of this project was not to evaluate the efficiency of the expression system used or to investigate other options, but to obtain milligram quantities of ^{15}N-labeled Long-[Leu^{60}]-IGF-I for NMR studies.

It was initially hoped that both ^{15}N-labeled Long-[Leu^{60}]-IGF-I and ^{13}C/^{15}N-labeled Long-[Leu^{60}]-IGF-I could be obtained, depending on the time available. Only one labeled

IGF sample had previously been expressed and purified in this laboratory, ^{15}N-labeled Long-[Arg^3]-IGF-I. The major problem with the production of ^{13}C-labeled protein was that 30 g/L glucose is normally used in the fermentation using the protocol outlined above. Trial fermentations were performed using low glucose media, however, very low optical density measurements were reached, in the range of 0.8, prior to glucose exhaustion. Given the time constraints and the expense of ^{13}C-glucose, only ^{15}N labeled protein was produced. Yields of up to 1 g fusion protein/l of fermentation broth are achieved using this expression system for the production of unlabelled proteins where nitrogen is not limited. With production of a ^{15}N-labeled protein, nitrogen is limited and the yield of inclusion bodies is lower. However, milligram quantities of Long-[Leu^{60}]-IGF-I were obtained, sufficient material for NMR spectroscopic studies.

CHAPTER 3

Resonance Assignments

3.1 Introduction

The sequential assignment method (Wüthrich et al., 1982) has been described in section 1.9.2. Obtaining sequential assignments is the essential first step toward three-dimensional structure determination for any protein using NMR. Each resonance must be associated with a specific nucleus in the protein. Resonance assignments must be sequence specific, that is, each resonance must be assigned to a spin in a particular amino acid residue in the protein sequence. These can be deduced from through-bond interactions (via scalar couplings), through-space interactions (via dipolar coupling), and chemical environment (via the chemical shift).

The combination of 2D ^{15}N-HSQC, 3D ^{15}N-TOCSY-HSQC and 3D ^{15}N-NOESY-HSQC (Zhang et al., 1994) experiments provides most of the through-bond and through-space connectivities necessary for sequential assignment of ^{15}N-labeled proteins. This method of sequential assignment for ^{15}N-labeled proteins was pioneered by Marion et al. (1989) and subsequently further developed by Ikura et al. (1990). Since proline residues do not have a backbone amide proton, neither the assignment of the proline spin systems or their connectivities to neighbouring residues can be established using these heteronuclear experiments alone. Homonuclear 2D NOESY and 2D TOCSY spectra (Oschkinat et al., 1989) are necessary for the assignment of proline proton resonances. Heteronuclear 2D ^{15}N-HSQC-TOCSY, 2D ^{15}N-HSQC-NOESY (Zhang et al., 1994) and 3D ^{15}N-HSQC-NOEY-HSQC (Zhang et al., 1994) experiments may also be used to resolve ambiguities remaining after complete analysis of the basic experiments. Resonance assignment, despite the advent of automated procedures, is still frequently the rate determining step of the three-dimensional structure determination. The determination of resonance assignments for Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I and the extent of these assignments are described in this chapter.

3.2 Materials and Hardware

Resonance assignment and the structure determination of Long-[Arg^3]-IGF-I were achieved with two samples, unlabelled Long-[Arg^3]-IGF-I at a concentration of 1.3 mM in 100% D_2O at pH 4.1 and ^{15}N-labeled Long-[Arg^3]-IGF-I at a concentration of 1.0 mM in 90% H_2O/10% D_2O at pH 3.0. One sample of Long-[Leu^{60}]-IGF-I was used, i.e. ^{15}N-

labeled Long-[Leu^{60}]-IGF-I at a concentration of 1.0 mM in 90% H_2O/10% D_2O at pH 3.0. The preparation of these samples was described in Chapter 2.

Preliminary NMR experiments on the Long-[Arg^{3}]-IGF-I sample were performed on a Varian VXR-500S NMR spectrometer. Subsequent NMR experiments were performed on a Varian INOVA-500 spectrometer operating at 500.058 MHz and 50.676 MHz for ^{1}H and ^{15}N respectively and a Varian INOVA-600 spectrometer operating at 600.046 MHz and 60.809 MHz for ^{1}H and ^{15}N respectively. All NMR experiments on Long-[Leu^{60}]-IGF-I were performed on a Varian INOVA-600 spectrometer.

3.3 Methods

3.3.1 Heteronuclear NMR Experiments

The following experiments were collected on the two ^{15}N-labeled proteins: a 2D ^{15}N-HSQC (Kay et al., 1992), a 3D sensitivity-enhanced ^{15}N-TOCSY-HSQC (Zhang et al., 1994) with a mixing time of 70 ms, two 3D sensitivity-enhanced ^{15}N-NOESY-HSQC (Zhang et al., 1994) experiments with mixing times of 80 ms and 140 ms, a 2D ^{15}N-HSQC-TOCSY (Zhang et al., 1994) with a mixing time of 70 ms, 2D ^{15}N-HSQC-NOESY (Zhang et al., 1994) with a mixing time of 100 ms, a 3D sensitivity-enhanced ^{15}N-HSQC-NOESY-HSQC (Frenkiel et al., 1990) with a mixing time of 140 ms, and a 3D sensitivity-enhanced ^{15}N-HSQC-TOCSY-NOESY-HSQC (Zhang et al., 1994) with a NOESY mixing time of 100 ms and a TOCSY mixing time of 35 ms. Pulse sequences for the NMR experiments are shown in Appendix 1.

The standard ^{15}N-separated experiments (outlined above) produced approximately 80% of the non-prolyl backbone assignments. The final 20% of the assignments were made with the 3D ^{15}N-HSQC-NOESY-HSQC and 3D ^{15}N-HSQC-TOCSY-NOESY-HSQC experiments. Dipolar couplings between two sequential amide protons, for example, HN(i) and HN(i+1) with the same H^N shift, were identified by spreading the signals according to the nitrogen shift of N(i+1) with the 3D ^{15}N-NOESY-HSQC experiment. However, if HN(i) and HN(i+1) themselves overlap, the 3D ^{15}N-NOESY-HSQC experiment cannot resolve this NOE interaction between them. The 3D ^{15}N-HSQC-NOESY-HSQC and 3D ^{15}N-HSQC-TOCSY-NOESY-HSQC experiments resolve NOEs between overlapping amide proton signals so long as the amide nitrogen shifts are not also degenerate (Pollock

et al., 1996). In each of these experiments, the ^{15}N chemical shift of the nitrogen, which is one-bond correlated to the amide proton where magnetisation originates, is correlated with the ^{1}H and ^{15}N chemical shifts of the destination amide (Zhang et al., 1994). The 3D ^{15}N-HSQC-NOESY-HSQC experiment is highly sensitive to helical regions since it relies on strong dipolar coupling between amide protons. By appending a TOCSY element to the NOESY element, the 3D ^{15}N-HSQC-TOCSY-NOESY-HSQC experiment can be used to detect random sequences and β-structures which have efficient TOCSY transfers.

The heteronuclear 2D spectra were acquired using ^{1}H and ^{15}N sweep widths of 7002.8 Hz and 1700 Hz respectively. Spectra were recorded with either 160, 200 or 256 t_1 increments and 2048 complex points in t_2. All 3D heteronuclear spectra were acquired using ^{1}H and ^{15}N sweep widths of 6489.3 Hz and 1700 Hz respectively. The hypercomplex data were collected over 128 complex x 48 complex x 512 complex points in the F_1, F_2 (^{15}N), and F_3 dimensions respectively, with the exception of the 3D ^{15}N-HSQC-TOCSY-NOESY-HSQC which was recorded with a data size of 90, 80 and 512 complex points in F_1, F_2 and F_3 respectively. A ^{15}N-filtered NOESY spectrum (Lee et al., 1994) with 80 ms mixing time was acquired for Long-[Leu^{60}]-IGF-I to obtain ^{1}H assignments for proline residues. Information obtained from the ^{15}N-filtered NOESY experiment was supplemented by a 2D NOESY with a mixing time of 120 ms on an unlabelled sample of [Leu^{60}]-IGF-I.

3.3.2 Homonuclear NMR Experiments

The following experiments were collected on the unlabelled Long-[Arg^{3}]-IGF-I sample; a 2D NOESY with a mixing time of 150 ms and a 2D sensitivity-enhanced TOCSY with a mixing time of 80 ms. These experiments employed pulsed field gradients for artefact suppression and WATERGATE (Piotto et al., 1992; Sklenár et al., 1993) for solvent suppression. These 2D homonuclear spectra were acquired with a sweep width of 6002.4 Hz and were recorded with 350 complex t_1 increments. Each FID consisted of 2048 complex points in t_2.

3.3.3 Spectral Processing

All data were processed using VNMR software (versions 5.2 and 6.0) (Varian Associates, Palo Alto, CA). Typically, 3D experimental data were processed with a low frequency solvent filter and a 90^{o} shifted sine bell in all three dimensions. Linear prediction was used

to extend the heteronuclear and homonuclear dimensions to improve their digital resolution (Olejniczak & Eaton, 1990). Typically, linear prediction was used to extend 48 complex points to 96 in the nitrogen dimension and to extend 128 complex points to 256 complex points in the proton dimension. The ^{15}N chemical shifts were referenced directly to liquid NH_3 at 0 ppm using the $^{15}N/^{1}H$ frequency ratio of 0.101329118 (Wishart et al., 1995b).

3.3.4 Spectral Analysis

Spectral analysis and peak picking were performed using the program XEASY, version 1.3.10 (Bartels et al., 1995). Firstly, each cross peak in the 2D ^{15}N-HSQC spectrum was labeled. This peak list was loaded into the 3D ^{15}N-TOCSY-HSQC spectrum which results in a list of peaks which lie on the diagonal of the proton planes. A strip was defined for each peak, the peak positions were adjusted and the strips were searched for additional overlapping peaks that could not be identified in the 2D ^{15}N-HSQC spectrum. Cross peaks resulting from scalar coupling connectivities were picked and assigned to the sidechain atoms of the spin system. Notes regarding the probable type of spin system were kept as a comment entry in the peak edit window. The TOCSY peak list produced was then loaded into the 3D ^{15}N-NOESY-HSQC spectrum and some peaks repositioned in the NOESY spectrum for which positions differed slightly in the NOESY and TOCSY spectra. Spectral correlations between the strips were calculated starting with a strip that corresponded to a spin system of a known amino acid type, e.g., Gly or Ala and potential sequential neighbours of the strip were displayed. The displayed strips were inspected visually to identify the one that most likely corresponded to the sequential neighbour. When deciding on the most probable sequential neighbour, detailed comparisons of the line-shapes of the intra residual peaks and the potential sequential peaks were made. These probable sequential neighbours were recorded as a comment entry. This procedure was repeated until a subsequence of spin systems could be mapped to the amino acid sequence and a strip list containing the assigned subsequence was saved. This procedure was repeated until the entire sequence of spin systems could be mapped onto the amino acid sequence.

3.4 Results

3.4.1 Long-[Arg^3]-IGF-I

Identification of the spin systems in Long-[Arg^3]-IGF-I was achieved by inspection of the scalar coupling connectivities within individual residues using the 2D ^{15}N-HSQC, 2D ^{15}N-HSQC-TOCSY and 3D TOCSY-HSQC spectra. Proline residues were assigned from the homonuclear sensitivity-enhanced 2D TOCSY spectrum. Sequential resonance assignment followed procedures pioneered by Wüthrich and coworkers (1982) and utilised 3D NOESY-HSQC, 3D ^{15}N-HSQC-NOESY-HSQC and 3D ^{15}N-HSQC-TOCSY-NOESY-HSQC spectra (Marion et al., 1989; Ikura et al., 1990). Connectivities from proline residues were assigned from analysis of the 2D NOESY spectra (NOEs from δ-CH_2 protons to H^{α} of the preceding residue). A region of the ^{1}H-^{15}N sensitivity-enhanced 2D ^{15}N-HSQC spectrum of Long-[Arg^3]-IGF-I is presented in Figure 3.1 and the ^{1}H and ^{15}N resonance assignments are given in Appendix 2. Of the non-prolyl ^{1}H and ^{15}N backbone resonances, 100% were assigned, while 92% of ^{1}H sidechain resonance assignments were made. The ^{1}H and ^{15}N chemical shifts for Long-[Arg^3]-IGF-I have been deposited in the BioMagResBank under BMRB accession number 4069 (http://www.bmrb.wisc.edu).

As well as the 76 major cross peaks in Figure 3.1, there is a subset of low intensity cross peaks that probably arise from a minor species in solution. None of these low intensity cross peaks gives rise to observable NOEs in the 3D NOESY-HSQC spectrum. Only one resonance, assigned to Gly^{42}, is split into two distinct peaks of approximately equal intensity. Inspection of the 3D NOESY-HSQC spectrum revealed an identical NOE pattern for each of these diagonal peaks. Gly^{42} is close to Pro^{39} and the doubling of its cross peaks may be a result of cis-trans isomerism at this proline. Both the Leu^{10} and the Val^{11} cross peak in the 2D ^{15}N-HSQC spectrum are broad and shifted upfield in the ^{1}H dimension. The Val^{11} cross peak is shifted upfield also in the native IGF-I (Sato et al., 1992) suggesting a similar structural element that is retained near the start of helix 1. However, the ^{1}H chemical shift of the amide proton of Leu^{10} is shifted upfield in Long-[Arg^3]-IGF-I in comparison to its position in IGF-I. This may be a result of the close proximity of the N-terminal extension to the N-terminal end of helix 1.

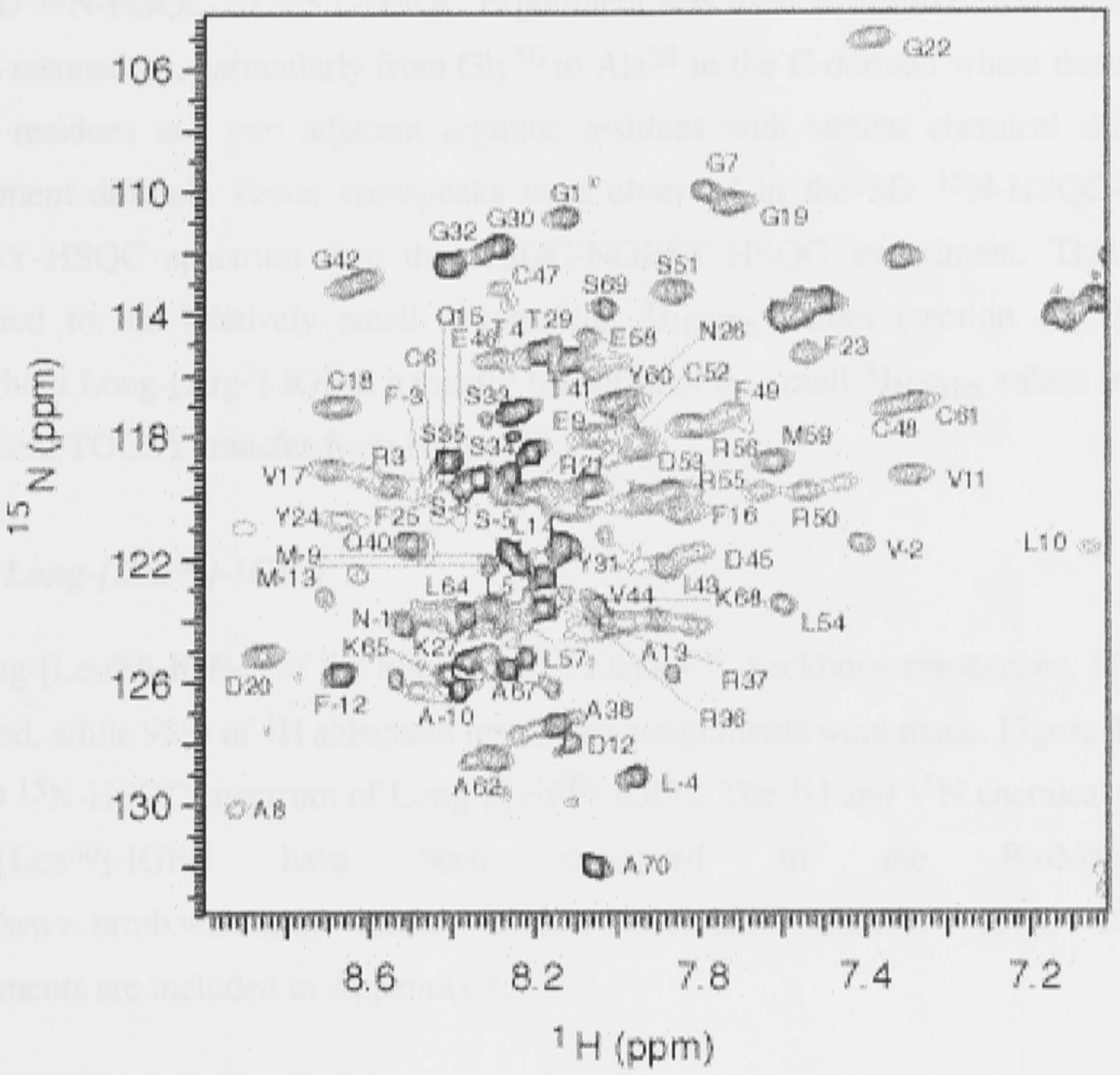

Figure 3.1: The 2D ^{15}N-HSQC spectrum of Long-[Arg^{3}]-IGF-I at 30 ^{o}C and pH 3.0. Assignments for the backbone amides are labeled by the amino acid one letter code followed by the sequence number.

The spin systems for the peptide segment Glu^{46} to Phe^{49} were difficult to assign as was the spin system of Cys^{52}. The segment Glu^{46} to Phe^{49} is encompassed by helix 2 (Cooke et al., 1991; Sato et al., 1992), so it is surprising that it is not easily assigned. However, Jansson et al.(1998) in their chemical shift mapping studies of IGF-I to IGFBP-1 also reported difficulties assigning the amino acid stretch of uncomplexed IGF-I from Asp^{45} to Phe^{49}, even at 50 ^{o}C.

Very few sequential NOEs were observed in the N-terminal extension, making this region quite difficult to assign, while intraresidue NOEs were lacking for Ser^{-5} and Leu^{-4}. Intraresidue NOEs were also not observed for Ala^{70}, the C-terminal residue of the protein.

The 3D ^{15}N-HSQC-NOESY-HSQC experiment was used to resolve overlapping amide proton resonances, particularly from Gly^{30} to Ala^{38} in the C-domain where three adjacent serine residues and two adjacent arginine residues with similar chemical shifts made assignment difficult. Fewer crosspeaks were observed in the 3D ^{15}N-HSQC-TOCSY-NOESY-HSQC spectrum than the HSQC-NOESY-HSQC experiment. This may be attributed to the relatively small intraresidue $^{3}J_{HNH\alpha}$ values (section 4.4.1) present throughout Long-[Arg^{3}]-IGF-I, a mostly helical protein. Small $^{3}J_{HNH\alpha}$ values resulted in inefficient TOCSY transfer from H^{N} to H^{α}.

3.4.2 Long-[Leu^{60}]-IGF-I

In Long-[Leu^{60}]-IGF-I, of the non-prolyl ^{1}H and ^{15}N backbone resonances, 100% were assigned, while 92% of ^{1}H sidechain resonance assignments were made. Figure 3.2 shows the 2D ^{15}N-HSQC spectrum of Long-[Leu^{60}]-IGF-I. The ^{1}H and ^{15}N chemical shifts for Long-[Leu^{60}]-IGF-I have been deposited in the BioMagResBank (http://www.brmb.wisc.edu) under BMRB accession number 4278. Resonance assignments are included in Appendix 3.

As with Long-[Arg^{3}]-IGF-I, resonances from the peptide segment Glu^{46} to Phe^{49} and Cys^{52} were difficult to assign though more difficulty was encountered in the assignment of Long-[Leu^{60}]-IGF-I. It is possible that the mutation at position 60 of Long-[Leu^{60}]-IGF-I, which is close in space to Glu^{46}-Phe^{49} in α-helix 2, disrupts the structure as compared with the parent protein and Long-[Arg^{3}]-IGF-I. Very few sequential NOEs were observed in either the N-terminal extension or the C-terminus and intra residue NOEs were not observed for some residues, namely Ser^{-5}, Leu^{-4} and Ala^{70}. These limited NOE data made assignment of both the N- and C-termini difficult, however assignments obtained for Long-[Arg^{3}]-IGF-I could be used as a guide. No additional low intensity cross peaks were present in the 2D ^{15}N-HSQC spectrum of Long-[Leu^{60}]-IGF-I (Figure 3.2) as were observed for Long-[Arg^{3}]-IGF-I.

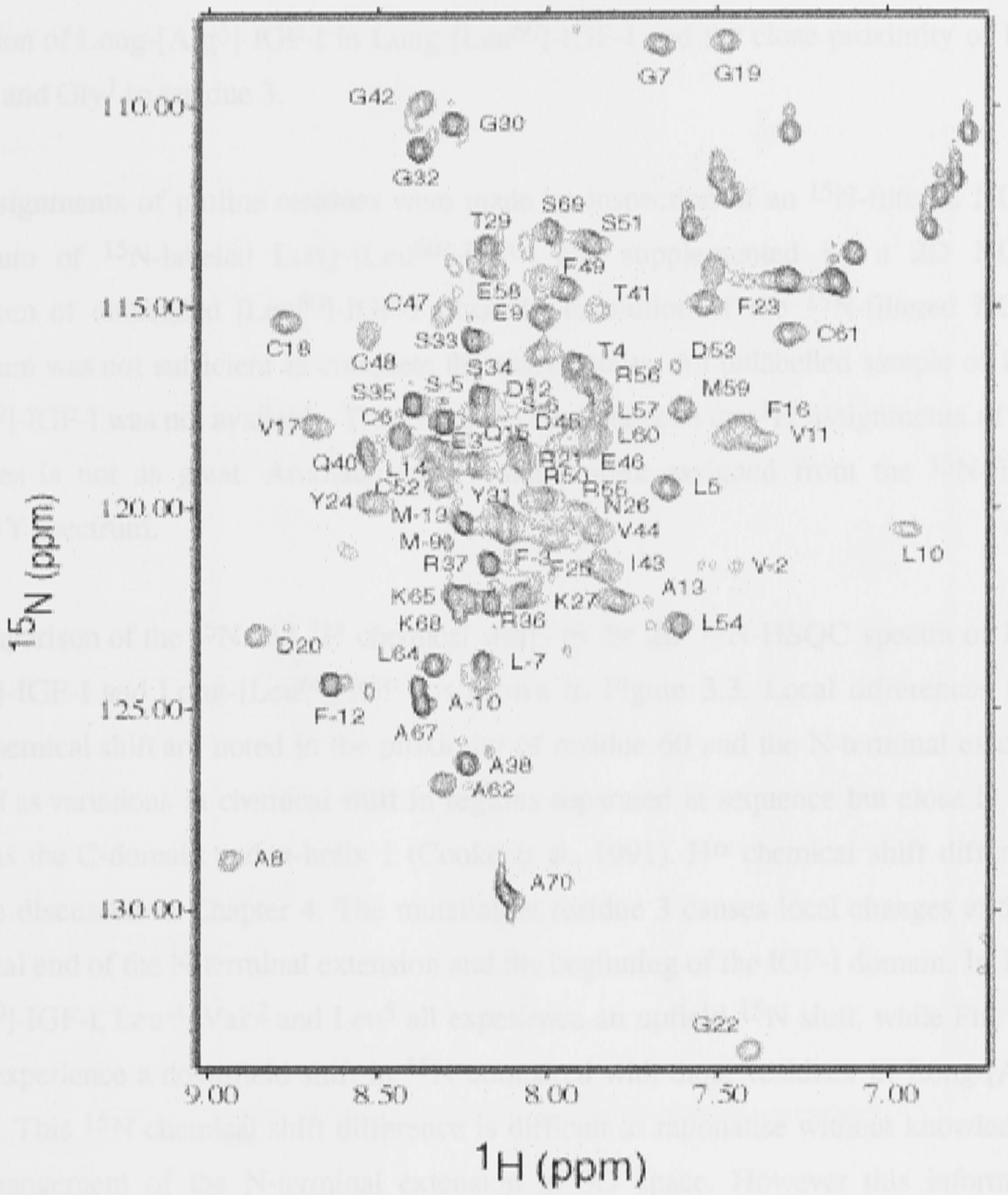

Figure 3.2: 2D ^{15}N-HSQC spectrum of 1 mM ^{15}N Long-[Leu^{60}]-IGF-I at 30 ^{o}C and pH 3.4 in 90% H_2O/D_2O. The assignments are labeled by the amino acid one letter code followed by the sequence number.

At 30 ^{o}C, no amide proton cross peaks were observed in the 2D ^{15}N-HSQC spectrum for Leu^{-4}, Asn^{-1} or Gly^{1}. However, the amide proton cross peak of Leu^{-4} was observed in the 3D NOESY-HSQC spectrum at this temperature. Amide proton cross peaks for all three residues were observed in the 2D ^{15}N-HSQC spectrum at 20 ^{o}C. This was not the case with Long-[Arg^{3}]-IGF-I where amide proton cross peaks for all three residues were observed in HSQC spectra acquired at both 20 ^{o}C and 30 ^{o}C. This difference suggests that these three residues are undergoing exchange more rapidly in Long-[Leu^{60}]-IGF-I than in Long-[Arg^{3}]-IGF-I which may be a result of the reversal of the Glu^{3} to Arg^{3}

mutation of Long-[Arg^3]-IGF-I in Long-[Leu^{60}]-IGF-I and the close proximity of Leu^{-4}, Asn^{-1} and Gly^1 to residue 3.

^1H assignments of proline residues were made by inspection of an ^{15}N-filtered NOESY spectrum of ^{15}N-labeled Long-[Leu^{60}]-IGF-I and supplemented by a 2D NOESY spectrum of unlabelled [Leu^{60}]-IGF-I since the resolution of the ^{15}N-filtered NOESY spectrum was not sufficient to complete the assignments. An unlabelled sample of Long-[Leu^{60}]-IGF-I was not available. Therefore, the confidence in the ^1H assignments of these residues is not as great. Aromatic spin systems were assigned from the ^{15}N-filtered NOESY spectrum.

A comparison of the ^{15}N and ^1H chemical shifts in the 2D ^{15}N-HSQC spectra of Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I is shown in Figure 3.3. Local differences in the ^{15}N chemical shift are noted in the proximity of residue 60 and the N-terminal extension as well as variations in chemical shift in regions separated in sequence but close in space such as the C-domain and α-helix 1 (Cooke et al., 1991). H^α chemical shift differences will be discussed in Chapter 4. The mutation at residue 3 causes local changes at the C-terminal end of the N-terminal extension and the beginning of the IGF-I domain. In Long-[Leu^{60}]-IGF-I, Leu^{-4}, Val^{-2} and Leu^5 all experience an upfield ^{15}N shift, while Phe^{-3} and Thr^4 experience a downfield shift in ^{15}N compared with those residues in Long-[Arg^3]-IGF-I. This ^{15}N chemical shift difference is difficult to rationalise without knowledge of the arrangement of the N-terminal extension in 3D space. However this information, combined with the absence of cross peaks for Asn^{-1} and Gly^1 at 30 °C indicates that the N-terminal extension of Long-[Leu^{60}]-IGF-I is likely to have undergone a minor change in average conformation in comparison with Long-[Arg^3]-IGF-I.

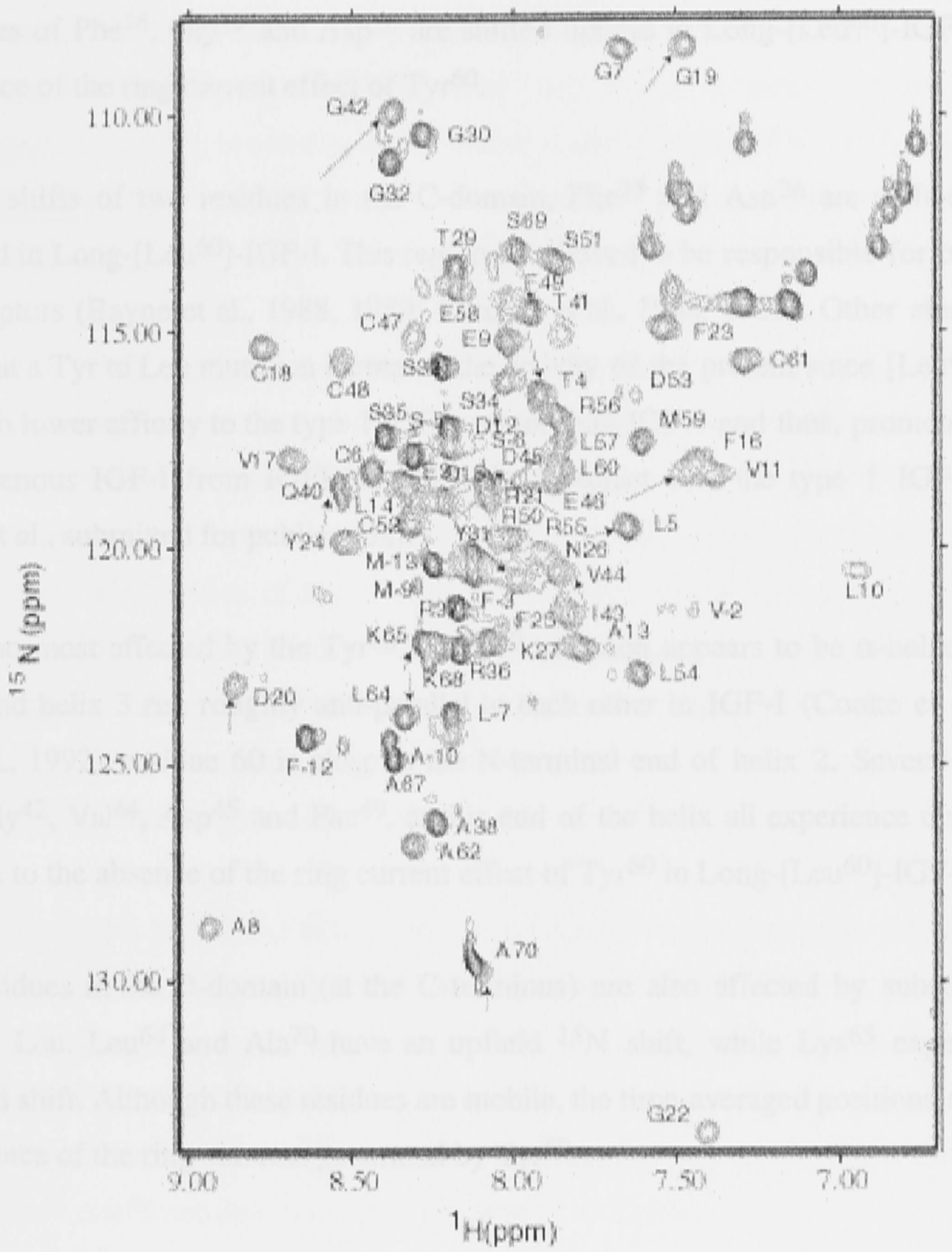

Figure 3.3: 2D ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled Long-[Leu^{60}]-IGF-I. Arrows indicate those resonances shifted by more than 1 ppm in the ^{15}N dimension when compared with the HSQC spectrum of Long-[Arg^{3}]-IGF-I (Figure 3.1). The origin point of the arrow approximates the resonance position of the residue in the spectrum of Long-[Arg^{3}]-IGF-I (Figure 3.1).

From Figure 3.3 it is apparent that Phe^{16}, Gly^{19} and Asp^{20} are all shifted slightly upfield in Long-[Leu^{60}]-IGF-I compared to Long-[Arg^{3}]-IGF-I. In IGF-I, Phe^{16} is part of helix 1, while Gly^{19} and Asp^{20} form part of a β-turn following helix 1 (Sato et al., 1992). These residues are close in three dimensional space to residue 60 which is the C-terminal residue of helix 3 and the three helices enclose a hydrophobic core. Presumably, the ^{15}N

resonances of Phe^{16}, Gly^{19} and Asp^{20} are shifted upfield in Long-[Leu^{60}]-IGF-I due to the absence of the ring current effect of Tyr^{60}.

The ^{15}N shifts of two residues in the C-domain, Phe^{25} and Asn^{26} are shifted slightly downfield in Long-[Leu^{60}]-IGF-I. This region is believed to be responsible for binding to IGF receptors (Bayne et al., 1988, 1989; Cascieri et al., 1988, 1989). Other studies have shown that a Tyr to Leu mutation increases the activity of the protein since [Leu^{60}]-IGF-I binds with lower affinity to the type 1 IGF receptor than IGF-I and thus, promotes release of endogenous IGF-I from IGFBPs which then interact with the type 1 IGF receptor (Milner et al., submitted for publication).

The region most affected by the Tyr^{60} to Leu^{60} mutation appears to be α-helix 2. Since helix 2 and helix 3 run roughly anti-parallel to each other in IGF-I (Cooke et al., 1991; Sato et al., 1992), residue 60 is close to the N-terminal end of helix 2. Several residues, Gln^{40}, Gly^{42}, Val^{44}, Asp^{45} and Phe^{49}, at this end of the helix all experience upfield ^{15}N shifts due to the absence of the ring current effect of Tyr^{60} in Long-[Leu^{60}]-IGF-I.

Three residues in the D-domain (at the C-terminus) are also affected by substitution of Tyr^{60} by Leu. Leu^{64} and Ala^{70} have an upfield ^{15}N shift, while Lys^{65} experiences a downfield shift. Although these residues are mobile, the time-averaged positions are within the influence of the ring currents generated by Tyr^{60}.

3.5 Discussion

In the first solution structure determination of IGF-I by Cooke et al. (1991), resonance assignments were made for 87% of the backbone resonances and 72% of the proton resonances. Assignments could not be made for either the H^{N} or the H$^{\alpha}$ protons of Ser^{34}, Ser^{35}, Arg^{36}, Arg^{37}, Arg^{50}, Lys^{68} or Ser^{69}, the H^{N} protons of Gly^{1} or Lys^{27} or the H$^{\alpha}$ protons of Pro^{28} or Pro^{63}. All of these assignment ambiguities, with the exception of the H^{N} proton of Gly^{1}, were resolved by Sato et al. (1992) although two ambiguous assignments, the H^{N} protons of Lys^{68} and Ala^{70}, were noted. A complete set of resonance assignments were obtained for mini-IGF-I (De Wolf et al., 1996). All three studies employed homonuclear NMR techniques using unlabelled samples. Although the solution structure of a protein of 70 amino acids residues is achievable under favourable

circumstances, using only homonuclear techniques, the helical structure and flexibility of these proteins suggested that ^{15}N labelling will enable a more precise structure determination. Since the two analogues of IGF-I, Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I both have 83 amino acid residues, the limitations of homonuclear NMR and the helical and flexible nature of these proteins, necessitated ^{15}N labelling.

Heteronuclear NMR experiments of Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I enabled many of the ambiguities associated with resonance assignments of IGF-I, e.g., residues of the D-domain, to be resolved. Information obtained from the 3D ^{15}N-TOCSY-HSQC spectra was limited and some assignments were made using only the 3D ^{15}N-NOESY-HSQC spectra. Assignment of the 3D ^{15}N-TOCSY-HSQC spectra beyond H^{α} was not possible for many residues, mainly due to the helical nature of the protein resulting in small $^3J_{NH\alpha}$ couplings which reduce the efficiency of magnetisation transfer. The presence of 20% dimer in Long-[Arg^3]-IGF-I (see section 5.4) also has an adverse effect on resonance line width.

The 3D ^{15}N-HSQC-NOESY-HSQC experiment enabled resonances in the segment Ser^{33} to Arg^{37} of both Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I to be assigned. The residues in this segment could not be assigned by a combination of 3D ^{15}N-TOCSY-HSQC and 3D ^{15}N-NOESY-HSQC because of overlap of the H^N protons of Ser^{33} and Ser^{34} and very similar ^{15}N shifts. In addition, neither $\alpha N(i,i+1)$ nor $\beta N(i,i+1)$ NOEs between Ser^{33}, Ser^{34} and Ser^{35} could be identified.

The N-terminal extension and a segment in the A-domain, Glu^{46} to Phe^{49}, of both Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I were difficult regions to assign. In the resonance assignment of the N-terminal extension of both analogues, there was a paucity of sequential and medium range NOEs, and in Long-[Leu^{60}]-IGF-I, intraresidue NOEs were not present for Leu^{-4}. The C-terminal end of α-helix 2 (Cooke et al, 1991; and Sato et al., 1993) was also difficult to assign in the two IGF-I analogues. Interestingly, no particular difficulties with resonance assignments were noted in this region of IGF-I by either Cooke et al. (1991) or Sato et al. (1992). However, Jansson et al. (1998) could not make any assignments for the polypeptide segment Asp^{45}-Phe^{49}, neither could assignments be made for Gly^1 or Cys^6. The segment Leu^{-4} to Thr^4 in Long-[Leu^{60}]-IGF-I was more difficult to assign than the same segment in Long-[Arg^3]-IGF-I due to the absence of some resonances at 30 ^oC in this region (section 3.3.2). This suggests that there are some

differences in the conformation or flexibility of Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I in this region, probably induced by the Glu^3 to Arg^3 mutation. The peptide segment Glu^{46}-Phe^{49} was also more difficult to assign in Long-[Leu^{60}]-IGF-I than in Long-[Arg^3]-IGF-I, which can be attributed to the proximity of this segment to Leu^{60} as determined by the solution structure of IGF-I (Cooke et al., 1991: Sato et al., 1993).

Jansson et al. (1998) made resonance assignments of IGF-I at 50 ^{o}C and pH 3.4 while Sato et al. (1992) used spectra obtained at both 40 ^{o}C and 50 ^{o}C and pH 3.0 to make resonance assignments, reporting that the spectral quality of IGF-I at 30 ^{o}C was not of sufficient resolution and only partial assignments were obtained. Specifically, Jansson et al. (1998) reported that at 30 ^{o}C the only residues for which assignments could be made unambiguously were Pro^2-Glu^3, Gly^7, Gly^{19}, Pro^{28}-Thr^{41}, Ser^{51}, Met^{59} and Pro^{63}-Ala^{70}. The segments Pro^{28}-Thr^{41} and Pro^{63}-Ala^{70} are both unstructured regions of IGF-I (Cooke et al., 1991; Sato et al., 1992). This suggests that there are some differences in the conformation or flexibility of IGF-I and both Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I at this temperature and pH.

The NMR chemical shift of a given nucleus depends on its electronic environment. It is sensitive to many atomic or molecular effects including the local electron distribution, bond hybridisation states, proximity to polar groups, nearby aromatic rings and local magnetic anisotropies (Wishart, 1992). The sensitivity of the H$^{\alpha}$ proton chemical shifts to the secondary structure of proteins has been used to deduce structural information. An analysis of the H$^{\alpha}$ chemical shift data for Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I is presented in Chapter 4 where the secondary structure of these two proteins is presented.

Amide proton chemical shifts are much more sensitive to the experimental conditions, such as temperature and pH, than the H$^{\alpha}$ chemical shifts and the relationship between the amide proton chemical shift and hydrogen bond length has been established for amphipathic α-helices (Zhou et al., 1992). A comparison of the amide proton chemical shifts of Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I with IGF-I and mini-IGF-I indicates local changes in molecular conformation and environment. The amide proton resonances of the following residues in Long-[Arg^3]-IGF-I were shifted in relation to both IGF-I and mini-IGF-I by more than 0.3 ppm: Cys^6, Gly^7, Ala^8, Leu^{10}, Gln^{15}, Val^{17}, Glu^{46}, Cys^{48}, Cys^{52}, Leu^{54} and Leu^{57}. The amide proton resonances of the following residues in Long-[Leu^{60}]-IGF-I were shifted relative to both IGF-I and mini-IGF-I by more than 0.3 ppm;

Cys^{6}, Gly^{7}, Ala^{8}, Leu^{10}, Val^{17}, Cys^{48}, Cys^{52} and Leu^{54}. The most significant amide nitrogen resonance shift differences observed in both Long-[Arg^{3}]-IGF-I and Long-[Leu^{60}]-IGF-I in comparison with IGF-I occur at Leu^{10} (far up field at 6.88 ppm) and Ala^{8} (far downfield at 8.98 pm). These unusually large shifts are attributed to the ring current effects due to the proximity of Phe^{25} to Ala^{8} and Leu^{10} in Long-[Arg^{3}]-IGF-I relative to IGF-I. Interestingly, none of the amide protons of residues in unstructured regions of either Long-[Arg^{3}]-IGF-I or Long-[Leu^{60}]-IGF-I are shifted significantly relative to IGF-I. The residues of Long-[Arg^{3}]-IGF-I and Long-[Leu^{60}]-IGF-I whose amide protons experience shifts relative to IGF-I are located prior to helix 1, within helix 1, helix 2 and helix 3 and the link joining helices 2 and 3 in IGF-I. This suggests that there is some minor change in the conformation or the environment of the amide protons of residues of Long-[Arg^{3}]-IGF-I and Long-[Leu^{60}]-IGF-I which are in structured regions of IGF-I. There may even be some disruption of the secondary structure in these two IGF-I analogues.

Analysis of ^{15}N chemical shifts of Long-[Arg^{3}]-IGF-I and Long-[Leu^{60}]-IGF-I indicates that there may be differences in conformation in the regions Leu^{-4}-Thr^{4}, Phe^{16}, Gly^{19}-Asp^{20}, Phe^{25}-Asn^{26}, Gln^{40}-Phe^{49} and Leu^{64}-Ala^{70} between the two variants (discussed at length in chapters 4 and 5). A similar comparison is not possible between IGF-I and the Long-IGF-I analogues due to IGF-I and mini-IGF-I NMR studies being performed on unlabelled samples.

CHAPTER 4

Secondary Structure

4.1 Introduction

The secondary structural elements of proteins can be identified from the chemical shift of the α protons, the relative intensities of NOE connectivities between H^N, H^α and H^β protons and by three bond J-coupling information (Wüthrich, 1986). α-helices are characterized by strong NN(i,i+1) NOEs, weak αN(i,i+1) and NN(i,i+2) NOEs and the presence of αN(i,i+3) and αβ(i,i+3) NOEs. Regions of β-sheet or extended structure are characterized by strong αN(i,i+1) and weak NN(i,i+1) NOEs. A sequential region of $^3J_{HNH\alpha}$ values < 6 Hz is indicative of α-helices, while regions of β-sheet have $^3J_{HNH\alpha}$ values > 8 Hz. Antiparallel β-sheet regions have long-range $H^N(i)$-$H^N(j)$, $H^N(i)$-$H^\alpha(j)$ and $H^\alpha(i)$-$H^\alpha(j)$ crosspeaks, where i and j represent residues from different β-strands in the protein. Parallel β-sheet regions have weak interstrand H^α-H^N and H^N-H^N crosspeaks. Turn structures are characterized by medium-range αN(i,i+2) NOEs and weak NN(i,i+1) NOEs. The following NOE crosspeaks may be present in tight turns:), αN(i,i+3), αN(i,i+2), NN(i,i+2) and NN(i,i+1). Weak and strong αN(i,i+1) and NN(i,i+1) crosspeaks are present in type 1 β-turns while the preferred position for proline in tight turns is position two. Reverse turns or loops which reverse the direction of the polypeptide chain at the surfaces of molecules, for example, β-hairpins and β-turns, are prevalent in globular proteins and are frequently classified as a third type of secondary structure in addition to helix and sheet. Turns can also be identified by amide proton exchange patterns and backbone dihedral angles, which will be discussed in Chapter 5. The NOE patterns and scalar couplings observed for the different types of secondary structure are shown in Table 4.1.

Analysis of the deviation of the chemical shifts of the H^α protons from their random coil values can further support secondary structure information gained from NOE connectivities since H^α chemical shifts are strongly dependent on the character and nature of protein secondary structure (Wishart et al., 1992). The H^α proton experiences an upfield shift with respect to the random coil value when in α helical conformation and a downfield shift when in a β-sheet conformation.

Hydrogen-deuterium exchange may be used to locate amide protons that are hydrogen-bonded or inaccessible to solvent. In structured regions of proteins such as an α-helix or a β-sheet, the exchange of amide hydrogens with solvent can be greatly slowed due to

hydrogen bonding, even when the bond is at the surface of the protein. The exchange process in these structured regions requires hydrogen bond breakage which may occur in small concerted folding reactions (Englander, 1975). For ^{15}N-labeled proteins with slow exchange ($K_{ex} < 0.01\ s^{-1}$), the slowly exchanging amides are identified by rapidly exchanging the protein from H_2O into D_2O buffer solution and repeatedly acquiring 2D ^{1}H-^{15}N HSQC spectra to observe the decrease in amide proton resonance intensities with time. The observation of a slowly exchanging amide proton implies that the amide proton may be involved in a hydrogen bond, but does not identify the atoms acting as hydrogen-bond acceptors.

Table 4.1. Summary of NOE and scalar couplings characteristic of secondary structure types

	α-helix	**β-sheet**
$d_{NN}(i, i+1)$	strong	weak
$d_{\alpha N}(i, i+1)$	weak	strong
$d_{\beta N}(i, i+1)$	medium/strong	weak
$d_{\alpha N}(i, i+3)$	present	not present
$d_{\alpha\beta}(i, i+3)$	present	not present
$d_{\beta N}(i, i+3)$	present	not present
$d_{\alpha N}(i, i)$	weak in 3D TOCSY-HSQC spectra	strong in 3D TOCSY-HSQC spectra, weak in 3D NOESY-HSQC spectra
$^{3}J_{NH\alpha}$	< 6 Hz	> 7.5 Hz
H^{α} shifts	upfield of random coil shift	downfield of random coil shift

The secondary structure, determined by analysis of NOE contacts, J-coupling information, chemical shift index analysis and amide proton exchange rates for both Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I is presented in this chapter.

4.2 Materials

^{15}N-labeled Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I and unlabelled Long-[Arg^3]-IGF-I samples were used for the secondary structure determination as described in Chapter 3. NMR experiments were performed on a Varian INOVA-600 spectrometer operating at 600.046 MHz and 60.809 MHz for ^{1}H and ^{15}N respectively.

4.3 Methods

4.3.1 NMR Experiments

In addition to the suite of NMR experiments described in Chapter 3, two additional NMR experiments were performed to determine $^{3}J_{NH\alpha}$ values, the ^{15}N-HMQC-J (Szyperski et al., 1992) and HNHA (Vuister and Bax, 1993). $^{3}J_{NH\alpha}$ values are used as confirmation of the type of secondary structure as determined by NOE patterns (discussed in section 4.1). Pulse sequences are shown for these experiments in Appendix 1. The technique of non linear fitting of J-modulated ^{15}N-^{1}H HMQC spectra for the determination of J-couplings decreases in accuracy when the line-width becomes significantly larger than the J-coupling. This presents a problem for the measurement of small J couplings, such as typically encountered in predominantly α-helical proteins. The HNHA experiment which correlates intraresidue $^{1}H_N$, ^{15}N and H^{α} resonances, relies on quantitative analysis of the diagonal-peak to cross peak intensity ratio in a ^{15}N-separated H^{N}-H^{α} homonuclear J correlation experiment. This experiment correlates the intra residue $^{1}H^{N}$, ^{15}N, and $^{1}H^{\alpha}$ resonances (Vuister and Bax, 1993).

4.3.2 Analysis of NOE Contacts

The NOEs observed in 3D NOESY-HSQC spectra acquired with a mixing time of 120 ms were most extensively used in the determination of regions of secondary structure in Long-[Arg^{3}]-IGF-I and Long-[Leu^{60}]-IGF-I. 3D NOESY-HSQC spectra acquired with a mixing time of 80 ms were used to confirm these contacts and eliminate the possibility of NOEs arising as a result of spin diffusion. A homonuclear 2D NOESY experiment with mixing time 120 ms was used to determine the intensity of αβ(i,i+3) NOEs which are characteristic of α-helices.

Since the solution structure of IGF-I has been previously determined, the intensity of sequential and medium-range NOEs in the helical regions of the IGF-I domain were investigated first, followed by the less structured regions. The intensity of NOEs throughout the remainder of the protein was subsequently analysed.

4.3.3 Coupling Constant Determination

$^{3}J_{NH\alpha}$ coupling constants were estimated from the splittings in the ^{15}N-HMQC-J spectrum by J-deconvolution using an "in-house" software package. The active splitting

was removed by dividing the time-domain signals by sin ($\pi J^* t$), equivalent to deconvolution in the frequency domain with an antiphase doublet. A search algorithm was used to find the condition where the trial coupling $J^* = J_a$, where J_a is the active coupling constant. Only a one-dimensional search is involved since J_a appears in both dimensions. Passive splittings were then eliminated by an analogous procedure involving cos ($\pi J^* t$). $^3J_{NH\alpha}$ coupling constants were also measured from the HNHA experiment (Vuister & Bax, 1993). The cross-peak to diagonal-peak intensity ratios (I_{cross}/I_{diag}) provides a direct measure of the magnitude of $^3J_{NH\alpha}$ using the equation:

$$I_{cross}/I_{diag} = -\tan 2(2\pi J_{HH}\zeta) \qquad (4.1)$$

where $^3J_{NH\alpha}$ is simplified to J_{HH} and ζ is the pulse sequence delay, 12.6 ms in this case. $^3J_{NH\alpha}$ values were considered quantitative and accurate (Vuister and Bax, 1993).

4.3.4 Amide Proton Exchange Rates

Approximate amide exchange rates were measured by a modified saturation transfer method (Spera et al., 1991; Grzesiek & Bax, 1993) to ascertain whether amide proton exchange rates were measurable by persistence methods. In the saturation transfer method, the ratio of the peak intensities was taken between two 2D ^{15}N-HSQC spectra, collected with and without weak irradiation of the water peak during recovery (I_S and I_0 respectively). The smaller the I_S/I_0 ratio, the more rapid the rate of amide proton exchange.

Qualitative amide proton exchange rates were determined by acquiring a series of 2D ^{15}N-HSQC spectra. The first spectrum collected was on the ^{15}N-labeled samples in 90% H_2O/10% D_2O. The protein was then lyophilised and reconstituted in approximately 400 µl D_2O. The second spectrum was acquired as soon as possible (10-15 min) following dissolution in D_2O. Further 2D ^{15}N-HSQC spectra were acquired at 1, 4, 8, and 24 hours. All spectra were collected under the same conditions at a temperature of 20 °C, lower than the temperature of 30 °C used for other NMR measurements (Scholtz and Robertson, 1995). Amide proton exchange experiments are sensitive to temperature, and acquiring the spectra at 20 °C was an attempt to slow the exchange of rapidly exchanging protons and assist in their identification. The acquisition time for each spectrum was 35 minutes. Spectra were acquired using ^{1}H and ^{15}N sweep widths of 6489.3 Hz and 1800 Hz respectively. Spectra were processed using a 90° shifted sinebell in both dimensions with

Varian VNMR version 5.2 software (Varian Associates, Palo Alto, CA). The 2D ^{15}N-HSQC spectra, taken at various elapsed times, were overlaid for inspection of the amide protons for which cross peaks remained. Amide protons, for which a cross peak in the 2D ^{15}N-HSQC experiment in D_2O was still observed after 24 hours, were classed as slowly exchanging. Amide protons, for which a cross peak in the 2D ^{15}N-HSQC experiment in D_2O was still observed after 4 hours, were classed as undergoing intermediate exchange.

4.3.5 Chemical Shift Index Analysis

Wishart et al. (1992) have demonstrated that the chemical shift of the nuclei of the protein backbone can be used to identify regions of secondary structure. The chemical shift index (CSI) method is a technique for rapidly and accurately determining the type and location of secondary structures in proteins from H^{α} chemical shifts. As was discussed in section 4.1, the H^{α} proton experiences an upfield shift with respect to the random coil value when in α-helical conformation and a downfield shift when in β-sheet conformation. If the α-proton chemical shift for a specific residue in the protein was greater than the random coil H^{α} proton chemical shift by 0.1 ppm for that particular amino acid (Wishart et al., 1992), it was assigned the label "1". If the α-proton chemical shift for a particular residue was less than the random coil H^{α} chemical shift for that particular amino acid by 0.1 ppm, it was labeled "-1". If the α-proton chemical shift for a particular residue was within the given range for the random coil H^{α} shift it was labeled "0".

Any sequence of four or more "-1s" that are not interrupted by a "1" is in an α-helical conformation (Wishart et al., 1992). Likewise, any sequence of three or more "1s" not interrupted by a "-1" is in a β-strand conformation. All other regions are designated as coil or unstructured. A local "density" of non zero chemical shift indices which exceeds 70% is required when defining regions of helical or extended structure. Termination points (at either end) of helices or β-strands can often be recognised by the first appearance of chemical shift indices that are opposite to those of the corresponding secondary structure. Otherwise, the first appearance of two consecutive zero-valued chemical shift indices marks the termination point.

Databases with random coil H^{α} chemical shift values have been developed by Wishart et al. (1992) and subsequently refined by Wishart and Sykes (1994) and Wishart et al. (1995a). In the latter study, the random coil ^{1}H, ^{13}C and ^{15}N chemical shifts for the 20

common amino acids when followed by alanine or proline were measured in addition to the random coil chemical shifts for alanine and proline when preceded by one of the 20 common amino acids.

4.4 Results

4.4.1 Long-[Arg^3]-IGF-I

The secondary structure elements in Long-[Arg^3]-IGF-I were identified from the relative intensities of sequential and medium-range NOEs and J-coupling information and were confirmed by analysis of the deviation of the chemical shifts of the H^α protons from their random coil chemical shift values. A diagrammatic summary of these results and amide proton exchange rates is presented in Figure 4.1.

When compared to the solution structure of IGF-I (Cooke et al., 1991; Sato et al., 1993) it is apparent from an examination of the pattern of NOEs of Long-[Arg^3]-IGF-I that the N-terminal extension Met^{-13} to Asn^{-1} has very little effect on the secondary structure of the IGF-I domain (Gly^1 to Ala^{70}) (Figure 4.1). Medium to strong sequential NN(i,i+1) NOEs are observed for the segments Ala^8 to Phe^{23}, Ile^{43} to Arg^{50} and Leu^{54} to Ala^{62} and a series of medium-range αN(i,i+3) NOEs are detected for the segments Ala^8 to Cys^{18}, Ile^{43} to Arg^{50} and Leu^{54} to Met^{59}. Furthermore, some αN(i,i+4) NOEs are also observed in these regions (Figure 4.2). αβ(i,i+3) NOEs are also observed for the segments Glu^9 to Cys^{18}, Ile^{43} to Cys^{48} and Leu^{54} to Met^{59}. Such NOE patterns are typically found in helical regions of proteins and hence this places the three helices in the IGF-I domain in approximately the same location as in the parent IGF-I protein. Strong αN(i,i+1) NOEs are observed in the segments Thr^4 to Leu^5, Cys^6 to Gly^7, and Gly^{30} to Ala^{38}. These types of NOEs suggest that the region between Gly^{30} and Ala^{38} is in an extended conformation. Medium-range αN(i,i+2) NOEs are detected between Gly^{19} and Arg^{21} and between Asp^{20} and Gly^{22} indicating that the β-turn observed in this region of IGF-I is most likely preserved in Long-[Arg^3]-IGF-I.

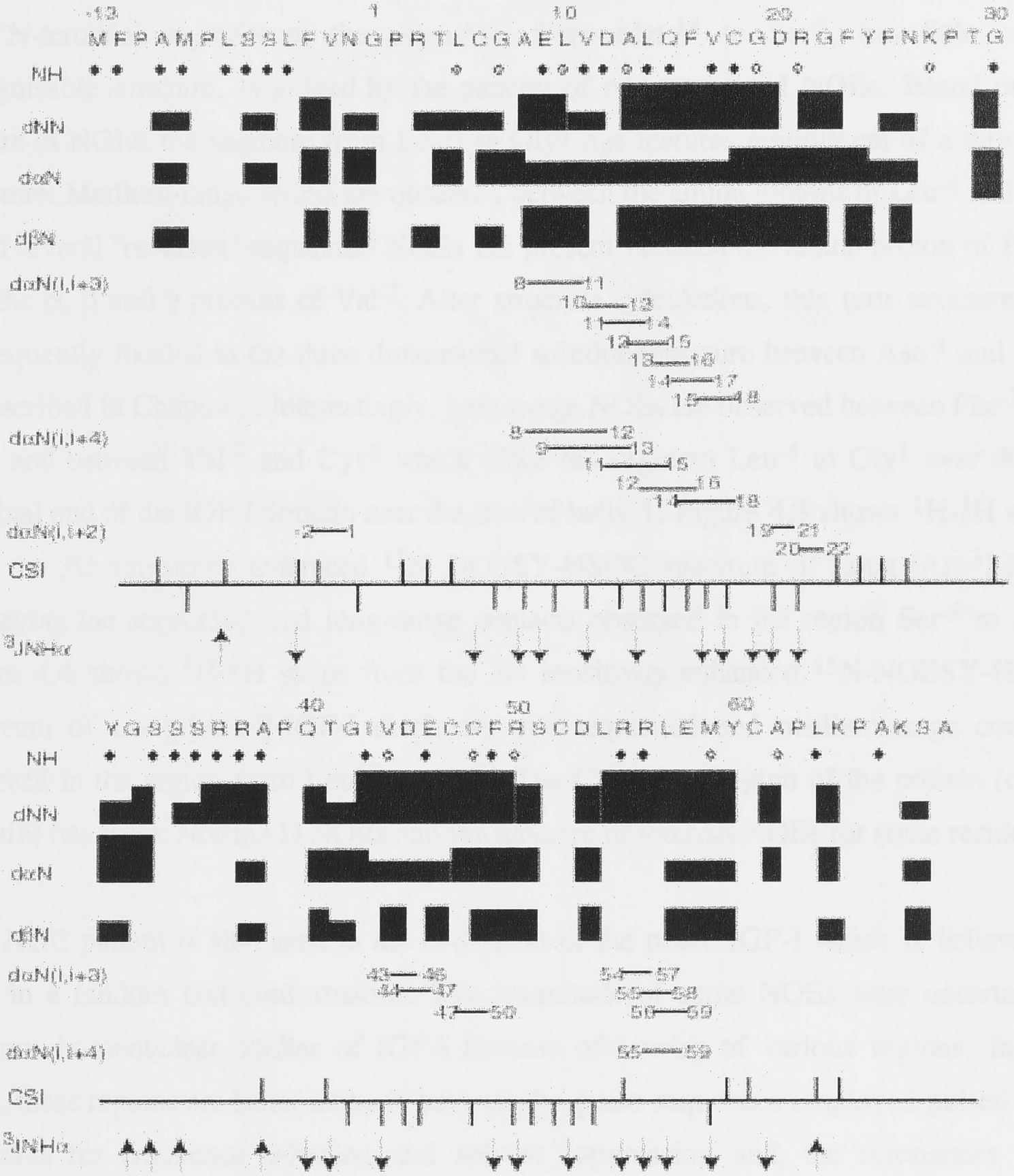

Figure 4.1: A summary of the NMR data defining the secondary structure of Long-[Arg^3]-IGF-I at pH 3.0 and 30 ^{o}C. The primary sequence is displayed at the top. Sequential NOE connectivities, d_{NN}, $d_{\alpha N}$ and $d_{\beta N}$, are indicated by bars linking the residues concerned and are classified as strong, medium or weak according to the thickness of the bar. Medium-range NOE connectivities, $d_{\alpha N}(i,i+2)$, $d_{\alpha N}(i,i+3)$ and $d_{\alpha N}(i,i+4)$ are linked by lines. The chemical shift index (defined as the deviation of the chemical shift value of the α proton from its random coil value) is classified as positive or negative by vertical lines above and below a baseline respectively. $^3J_{NH\alpha}$ values greater than 7.5 Hz are indicated by an upright arrow, $^3J_{NH\alpha}$ values less than 6 Hz are indicated by an inverted arrow. Unfilled circles indicate amide proton resonances that are observable 1 h after exchange of the solvent from H_2O to D_2O. Filled circles indicate amide proton resonances that are observed 24 h after solvent exchange.

The N-terminal segment of the extension, from Met^{-13} to Ser^{-5}, has little or no recognisable structure, as judged by the paucity of non-sequential NOEs. Based on the pattern of NOEs, the segment from Leu^{-4} to Gly^{1} has features reminiscent of a turn-type structure. Medium-range NOEs are observed between the amide protons of Leu^{-4} and Val-2 and several "reversed" sequential NOEs are present between the amide proton of Phe-3 and the α, β and γ protons of Val^{-2}. After structure calculations, this turn structure was subsequently located in the three dimensional solution structure between Asn^{-1} and Arg^{3} as described in Chapter 5. Interestingly, long-range NOEs are observed between Phe^{-3} and Gly^{7} and between Val^{-2} and Cys^{6} which place the segment Leu^{-4} to Gly^{1} over the N-terminal end of the IGF-I domain near the start of helix 1. Figure 4.3 shows ^{1}H-^{1}H strips from the 3D sensitivity enhanced ^{15}N NOESY-HSQC spectrum of Long-[Arg^{3}]-IGF-I displaying the sequential and long-range contacts observed in the region Ser^{-5} to Ala^{8}. Figure 4.4 shows ^{1}H-^{1}H strips from the 3D sensitivity enhanced ^{15}N-NOESY-HSQC spectrum of Long-[Arg^{3}]-IGF-I displaying the sequential and medium-range contacts observed in the region from Leu^{57} to Ala^{62}. The C-terminal region of the protein (or D-domain) has weak NN(i,i+1) NOEs and the absence of internal NOEs for some residues.

This NOE pattern is also seen in the D-domain of the native IGF-I which is believed to exist in a random coil conformation. The magnitude of some NOEs were uncertain in previous homonuclear studies of IGF-I because of overlap of various regions. In this work, these regions are better defined, because the pulse sequences employed pulsed field gradients for coherence selection and solvent suppression and, the resonances were dispersed along a third heteronuclear dimension (^{15}N). Rapidly exchanging protons are more readily observed because of these recent technological developments. Many of the long-range NOEs that are observed in native IGF-I are also observed in the IGF-I domain of Long-[Arg^{3}]-IGF-I, for example, Val^{11}-Val^{44}, Leu^{14}-Tyr^{60}, Leu^{14}-Phe^{23}, Phe^{16}-Leu^{54} and Ile^{43}-Tyr^{60}, suggesting that the addition of the N-terminal extension has only minor local effects on the three dimensional fold of the parent IGF-I domain.

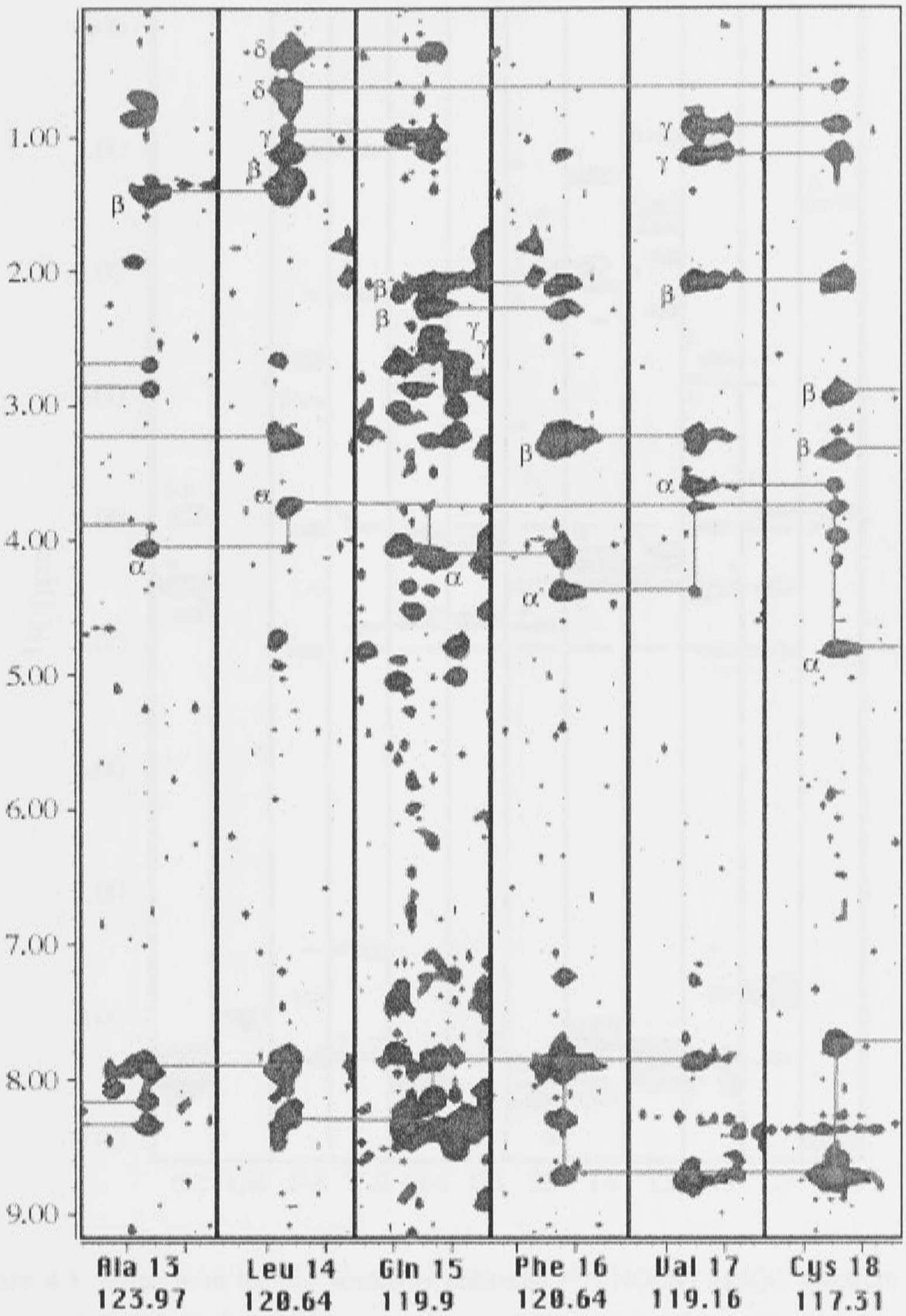

Figure 4.2: ^{1}H-^{1}H strips from the 3D sensitivity enhanced NOESY-HSQC spectrum of Long-[Arg^{3}]-IGF-I displaying the sequential and medium-range contacts for part of helix 1 encompassing residues Ala^{13} to Cys^{18}.

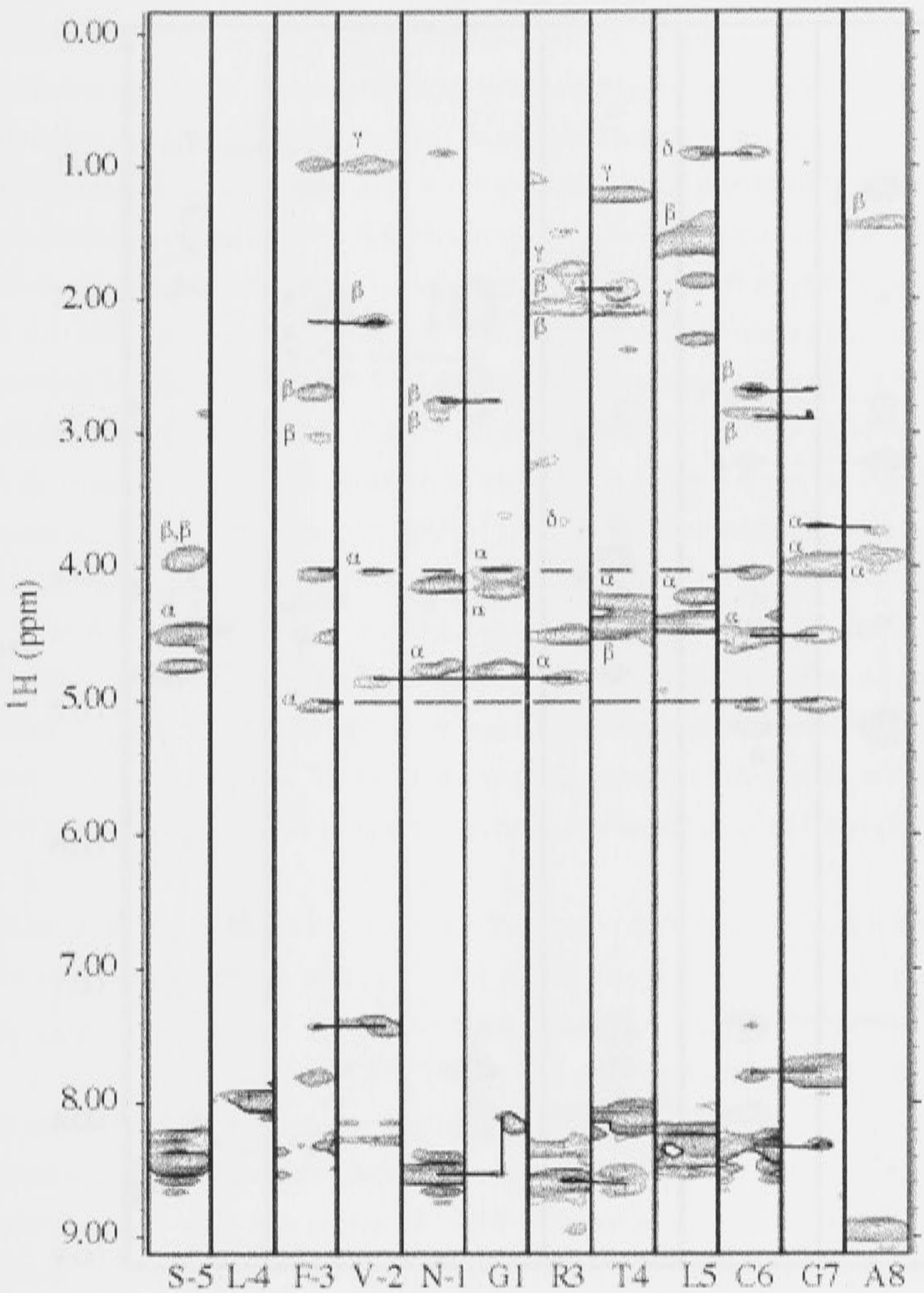

Figure 4.3: Strips from the 3D sensitivity enhanced ^{15}N-NOESY-HSQC spectrum of Long-[Arg^3]-IGF-I displaying the sequential and long-range contacts observed for the region Ser^{-5} to Ala^8. Long-range contacts are shown by dashed lines.

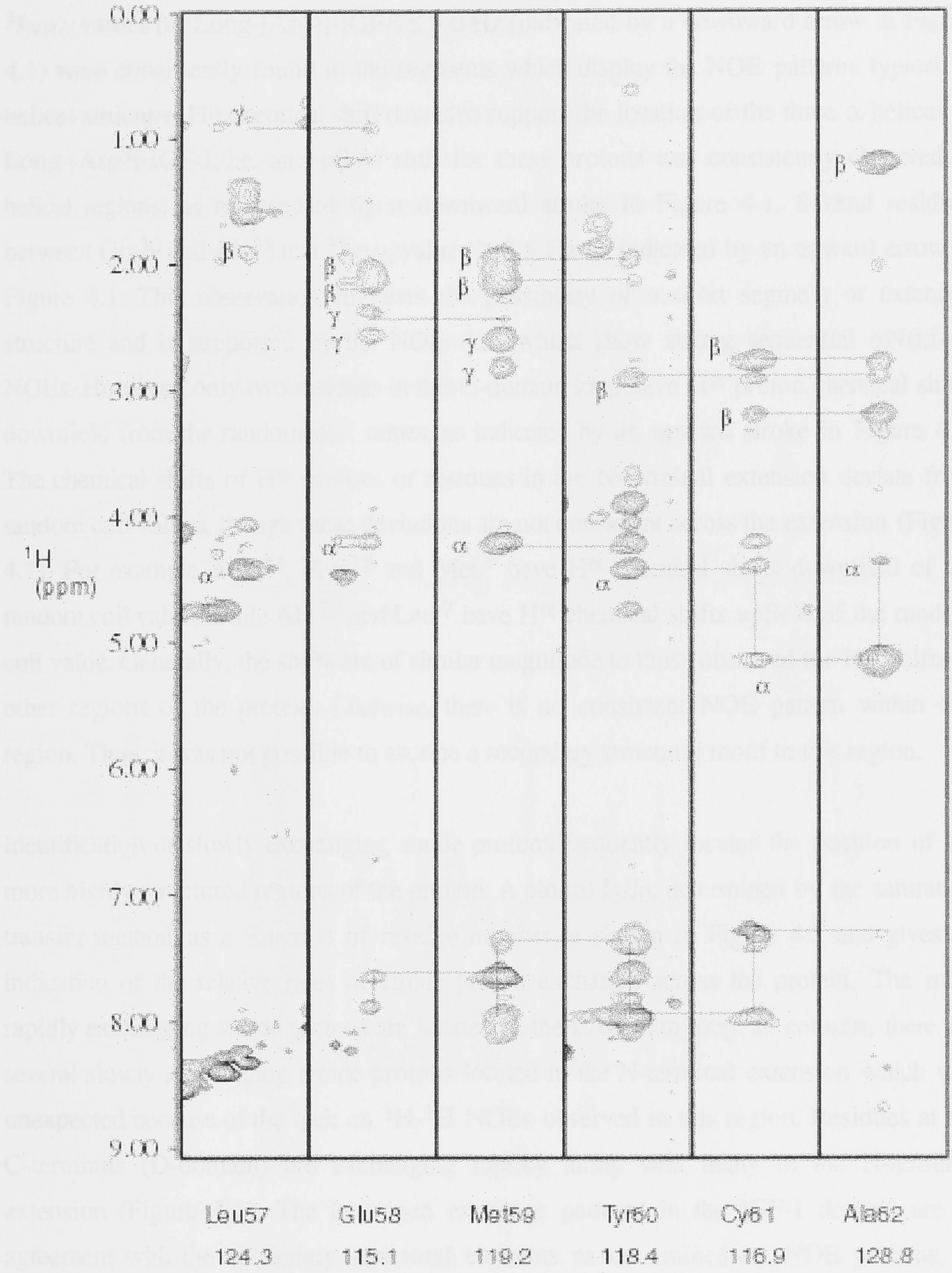

Figure 4.4: Strips from the 3D sensitivity enhanced ^{15}N-NOESY-HSQC spectrum of Long-[Arg^3]-IGF-I displaying the sequential and medium-range contacts observed for the region Leu^{57} to Ala^{62}.

$^3J_{NH\alpha}$ values for Long-[Arg^3]-IGF-I ≤ 5.0 Hz (indicated by a downward arrow in Figure 4.1) were consistently found in the segments which display the NOE patterns typical of helical structure. H^α chemical shift data also support the location of the three α-helices in Long-[Arg^3]-IGF-I, i.e. an upfield shift for these protons was consistently observed in helical regions, as represented by a downward stroke in Figure 4.1. Several residues between Gly^{30} and Ala^{38} had $^3J_{NH\alpha}$ values ≥ 7.5 Hz as indicated by an upward arrow in Figure 4.1. This observation suggests the possibility of a short segment of extended structure and is supported by the NOE data which show strong sequential αN(i,i+1) NOEs. However, only two residues in this C-domain loop have H^α proton chemical shifts downfield from the random coil values, as indicated by an upward stroke in Figure 4.1. The chemical shifts of H^α protons of residues in the N-terminal extension deviate from random coil values, though these deviations are not consistent across the extension (Figure 4.1). For example, Met^{-13}, Phe^{-12} and Met^{-9} have H^α chemical shifts downfield of the random coil values while Ala^{-10} and Leu^{-7} have H^α chemical shifts upfield of the random coil value. Generally, the shifts are of similar magnitude to those observed for H^α shifts in other regions of the protein. Likewise, there is no consistent NOE pattern within this region. Thus, it was not possible to ascribe a secondary structural motif to this region.

Identification of slowly exchanging amide protons frequently locates the position of the more highly structured regions of the protein. A plot of I_S/I_0, determined by the saturation transfer method, as a function of residue number is shown in Figure 4.5 and gives an indication of the relative rates of amide proton exchange across the protein. The most rapidly exchanging amide protons are located in the C-domain loop. In contrast, there are several slowly exchanging amide protons located in the N-terminal extension which was unexpected because of the lack on ^1H-^1H NOEs observed in this region. Residues at the C-terminus (D-domain) are exchanging rapidly along with many in the N-terminal extension (Figure 4.1). The hydrogen exchange patterns in the IGF-I domain are in agreement with the secondary structural elements as determined by NOE patterns, J-coupling data and supported by H^α chemical shift index analysis. Hydrogen exchange patterns in the IGF-I domain are also in agreement with previous homonuclear studies of wild-type IGF-I. Helix 1 shows consistent slow amide exchange along the whole length of the helix whereas only some of the protons within the two shorter helices are protected from the solvent. The only residues in helices 2 and 3 which have amide proton resonances still present 24 hours after sample preparation are Ile^{43}, Glu^{46} and Phe^{49} in helix 2, and Arg^{55} and Arg^{56} in helix 3. Residues within the C-domain loop (Gly^{30} to

Ala[38]) are also protected from solvent. This region in native IGF-I was poorly defined in the homonuclear studies and did not display any solvent protection (Sato et al., 1992), thus it is not clear whether this is a real difference between Long-[Arg[3]]-IGF-I and IGF-I or is due to a limitation of the homonuclear NMR methods used in IGF-I studies. IGF-II has a similar stretch of residues in an extended conformation between helix 1 and helix 2 that are also protected from solvent exchange, however, compared to Long-[Arg[3]]-IGF-I, this region is offset five residues toward the N-terminus, i.e. from Gly[25] to Phe[28] (Torres et al., 1995). Residues Gly[25] to Phe[28] are involved with high affinity binding to the type 1 IGF receptor (Cascieri et al., 1988, 1989).

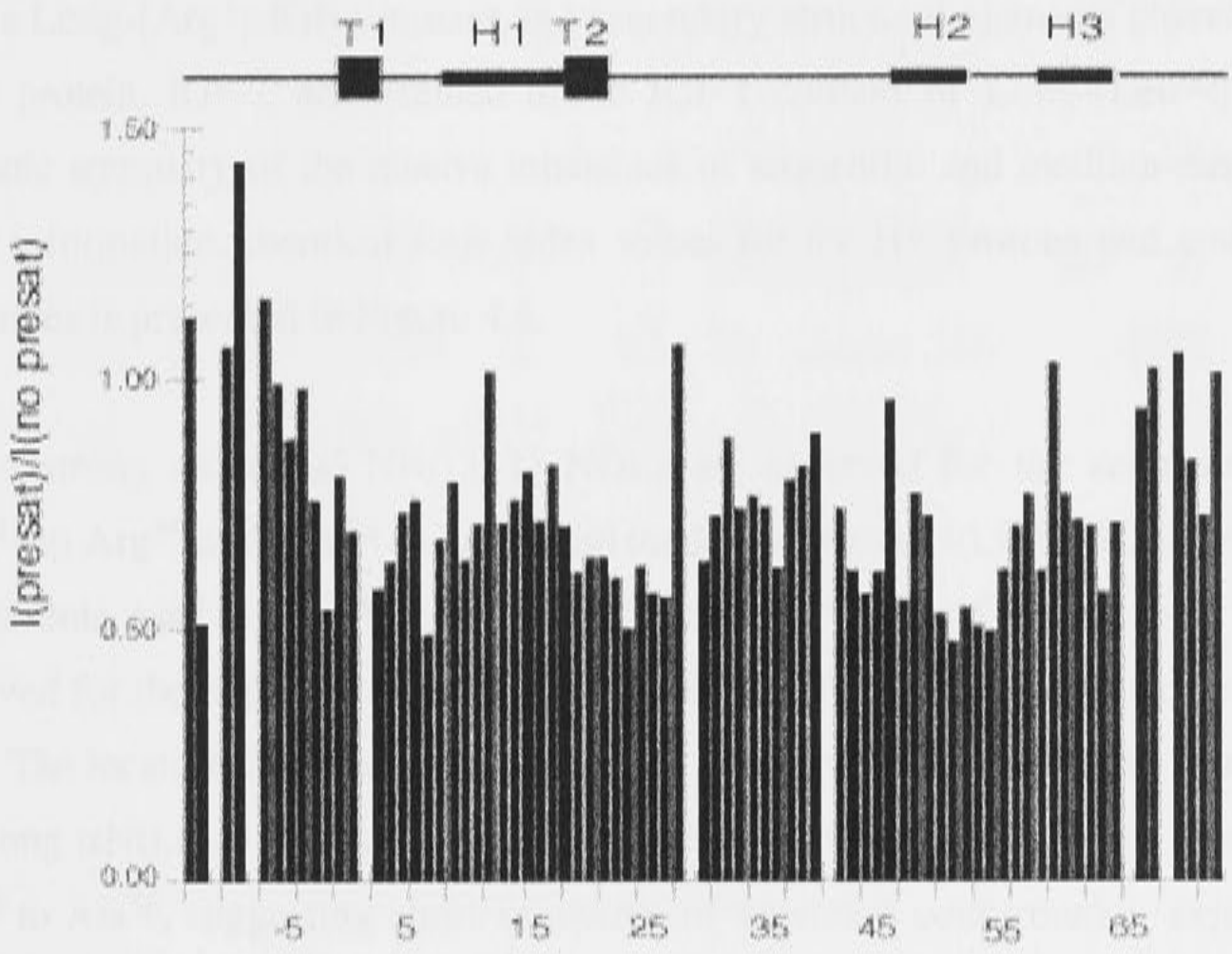

Figure 4.5: Plot of I_S / I_0 vs residue number which was used to estimate of the relative rates of amide proton relaxation.

A long stretch of residues in the N-terminal extension (Met[-13] to Leu[-4]) is also protected from solvent (Figure 4.1). This is a surprising result since generally, most protected amide protons are involved in hydrogen bonds (discussed in section 4.2), but no well defined secondary structure exists in this part of the N-terminal extension. The pattern of solvent

protection may reflect the hydrophobic nature of the extension which arranges the residues in such a way as to exclude solvent. However, these structures must be transitory since this region has no well defined secondary structure as indicated by the paucity of sequential, medium-range, or long-range NOE contacts between this part of the N-terminal extension and the IGF-I domain (Figure 4.5). Insight into what type of transitory structures exist may be gained from the three-dimensional solution structure of Long-[Arg^3]-IGF-I generated from the NOE data (discussed in Chapter 5). However, information pertaining to the structure of this segment will be limited if no further long-range contacts can be extracted from the NMR data following examination of the preliminary structures.

4.4.2 Long-[Leu^{60}]-IGF-I

As with the Long-[Arg^3]-IGF-I mutant, the secondary structural elements characteristic of the parent protein, IGF-I, are retained in the IGF-I domain of Long-[Leu^{60}]-IGF-I. A diagrammatic summary of the relative intensities of sequential and medium-range NOEs, J-coupling information, chemical shift index values for the H^{α} protons and amide proton exchange rates is presented in Figure 4.6.

Medium to strong sequential NN(i,i+1) NOEs are observed for the segments Ala^8 to Phe^{23}, Ile^{43} to Arg^{50} and Leu^{54} to Leu^{60} and medium-range αN(i,i+3) NOEs are detected for the segments Ala^8 to Cys^{18}, Ile^{43} to Arg^{50} and Leu^{54} to Leu^{60}. αN(i,i+4) NOEs are also observed for the region Ala^8 to Cys^{18} (Figure 4.7). These NOE patterns are indicative of helices. The location of these three helices are the same as in IGF-I and Long-[Arg^3]-IGF-I. Strong αN(i,i+1) NOEs are observed in the segments Glu^3 to Thr^4, Leu^5 to Gly^7 and Gly^{30} to Ala^{38}, suggesting short segments of extended conformation exist in these regions.

Figure 4.6: A summary of the NMR data defining the secondary structure of Long-[Leu^{60}]-IGF-I at pH 3.0 and 30 ^{o}C. The primary sequence is displayed at the top. Sequential NOE connectivities, d_{NN}, $d_{\alpha N}$ and $d_{\beta N}$, are indicated by bars linking the residues concerned and are classified as strong, medium or weak according to the thickness of the bar. Medium-range NOE connectivities, $d_{\alpha N}(i,i+2)$, $d_{\alpha N}(i,i+3)$ and $d_{\alpha N}(i,i+4)$ are linked by lines. The chemical shift index is classified as positive or negative by vertical lines above and below a baseline respectively. $^{3}J_{NH\alpha}$ values greater than 7.5 Hz are indicated by an upright arrow, $^{3}J_{NH\alpha}$ values less than 6 Hz are indicated by an inverted arrow. Open circles indicate amide proton resonances that are observable 1 h after exchange of the solvent from H_2O to D_2O. Filled circles indicate amide proton resonances that are observed 24 h after solvent exchange.

A medium-range αN(i,i+2) NOE was detected between Gly^{19} and Arg^{21}, which was also observed in Long-[Arg^3]-IGF-I. However, another medium-range αN(i,i+2) NOE was observed between Asp^{20} and Gly^{22} in Long-[Arg^3]-IGF-I which was not present in Long-[Leu^{60}]-IGF-I. This suggests that the β-turn that is present in this region for IGF-I and Long-[Arg^3]-IGF-I is preserved in Long-[Leu^{60}]-IGF-I but is not as well defined. There are differences in ^{15}N chemical shifts for residues Gly^{19} and Asp^{20} in Long-[Leu^{60}]-IGF-I compared with Long-[Arg^3]-IGF-I.

As with Long-[Arg^3]-IGF-I, Met^{-13} to Ser^{-5} of the N-terminal extension of Long-[Leu^{60}]-IGF-I has very few sequential NOEs and no non-sequential NOEs suggesting this segment has little or no recognisable structure. A weak NN(i,i+1) NOE was observed between Leu^{-4} and Phe^{-3} and the "reversed" sequential NOEs are observed between the amide proton of Phe^{-3} and the α, β and γ protons of Val^{-2} as in Long-[Arg^3]-IGF-I. Long-range NOEs between Phe^{-3} and Gly^7 and between Val^{-2} and Cys^6 are observed in Long-[Leu^{60}]-IGF-I as in Long-[Arg^3]-IGF-I, although there are some differences in the intensities of the NOEs between the two proteins. Figure 4.8 shows ^1H-^1H strips from the 3D ^{15}N NOESY-HSQC spectrum of Long-[Leu^{60}]-IGF-I displaying the sequential, medium- and long-range contacts observed in the region Ser^{-5} to Ala^8. The Arg^3 to Glu^3 mutation from Long-[Arg^3]-IGF-I to Long-[Leu^{60}]-IGF-I resulted in ^{15}N chemical shift differences of the amide protons between Leu^{-4} and Leu^5 indicating an altered conformation of this region in Long-[Leu^{60}]-IGF-I, however, some important elements appear to be conserved. Figure 4.9 shows the ^1H-^1H strips from the 3D ^{15}N-NOESY-HSQC spectrum of Long-[Leu^{60}]-IGF-I displaying the sequential and medium-range contacts observed in the region Leu^{57} to Ala^{62}, which are similar to those in Long-[Arg^3]-IGF-I (Figure 4.4). Only weak sequential NOEs are observed in the C-terminal region of Long-[Leu^{60}]-IGF-I and internal NOEs are absent for Ala^{70}, as observed in Long-[Arg^3]-IGF-I.

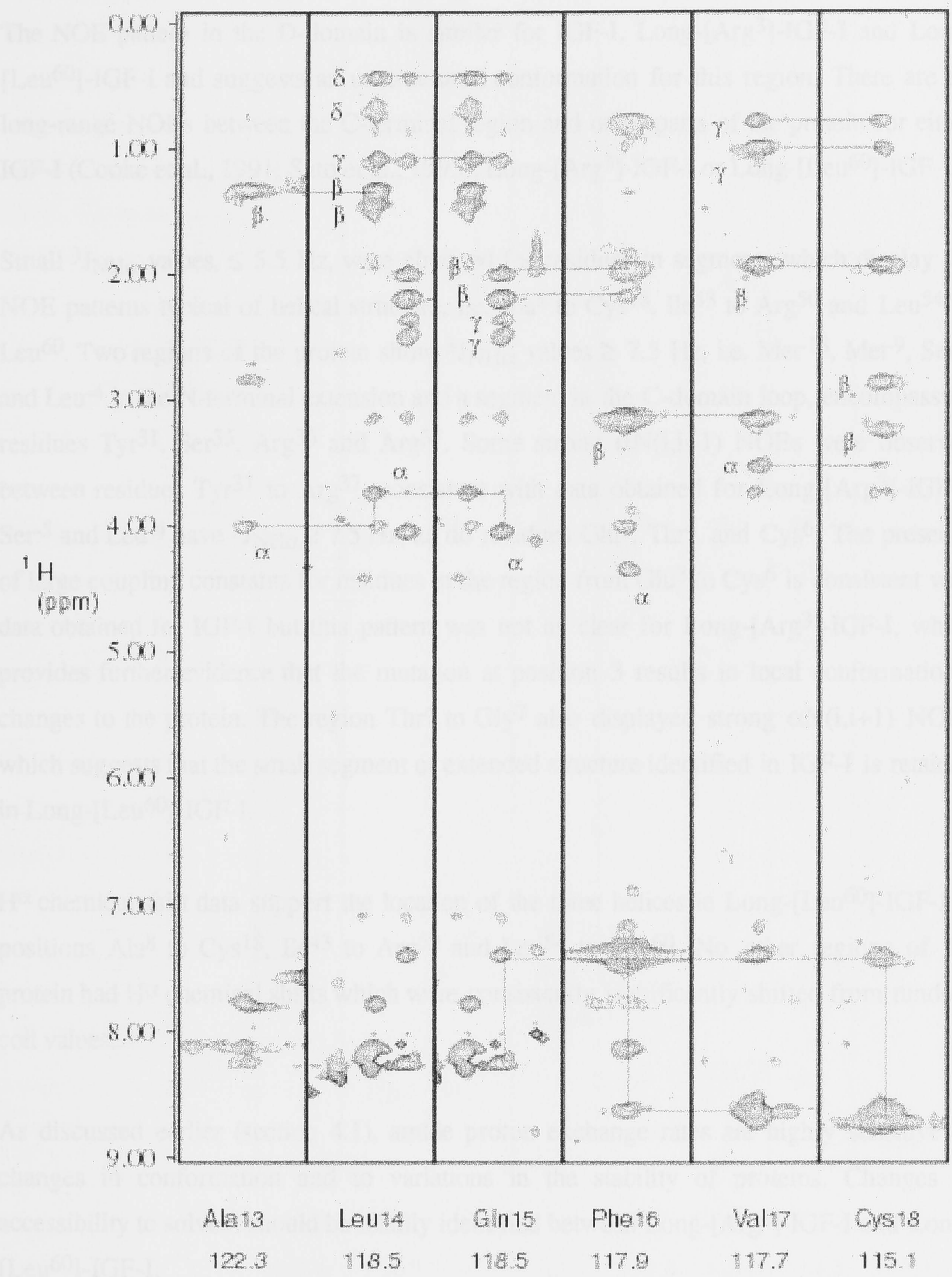

Figure 4.7: Strips from the 3D sensitivity enhanced ^{15}N-NOESY-HSQC spectrum of Long-[Leu^{60}]-IGF-I displaying the sequential and long-range contacts observed for the region Ala^{13} to Cys^{18}.

The NOE pattern in the D-domain is similar for IGF-I, Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I and suggests an unstructured conformation for this region. There are no long-range NOEs between the C-terminal region and other parts of the protein for either IGF-I (Cooke et al., 1991; Sato et al., 1993), Long-[Arg^3]-IGF-I or Long-[Leu^{60}]-IGF-I.

Small $^3J_{NH\alpha}$ values, ≤ 5.5 Hz, were obtained for residues in segments which display the NOE patterns typical of helical structure, i.e. Ala^8 to Cys^{18}, Ile^{43} to Arg^{50} and Leu^{54} to Leu^{60}. Two regions of the protein show $^3J_{NH\alpha}$ values ≥ 7.5 Hz, i.e. Met^{-13}, Met^{-9}, Ser^{-5} and Leu^{-4} in the N-terminal extension and a segment in the C-domain loop, encompassing residues Tyr^{31}, Ser^{35}, Arg^{36} and Arg^{37}. Some strong αN(i,i+1) NOEs were observed between residues Tyr^{31} to Arg^{37}, consistent with data obtained for Long-[Arg^3]-IGF-I. Ser^{-5} and Leu^{-4} have $^3J_{NH\alpha} \geq 7.5$ Hz, as do residues Glu^3, Thr^4, and Cys^6. The presence of large coupling constants for residues in the region from Glu^3 to Cys^6 is consistent with data obtained for IGF-I but this pattern was not as clear for Long-[Arg^3]-IGF-I, which provides further evidence that the mutation at position 3 results in local conformational changes to the protein. The region Thr^4 to Gly^7 also displayed strong αN(i,i+1) NOEs which suggests that the small segment of extended structure identified in IGF-I is retained in Long-[Leu^{60}]-IGF-I.

H^α chemical shift data support the location of the three helices in Long-[Leu^{60}]-IGF-I at positions Ala^8 to Cys^{18}, Ile^{43} to Arg^{50} and Leu^{54} to Leu^{60}. No other regions of the protein had H^α chemical shifts which were consistently significantly shifted from random coil values.

As discussed earlier (section 4.1), amide proton exchange rates are highly sensitive to changes in conformation and to variations in the stability of proteins. Changes in accessibility to solvent should be readily identified between Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I.

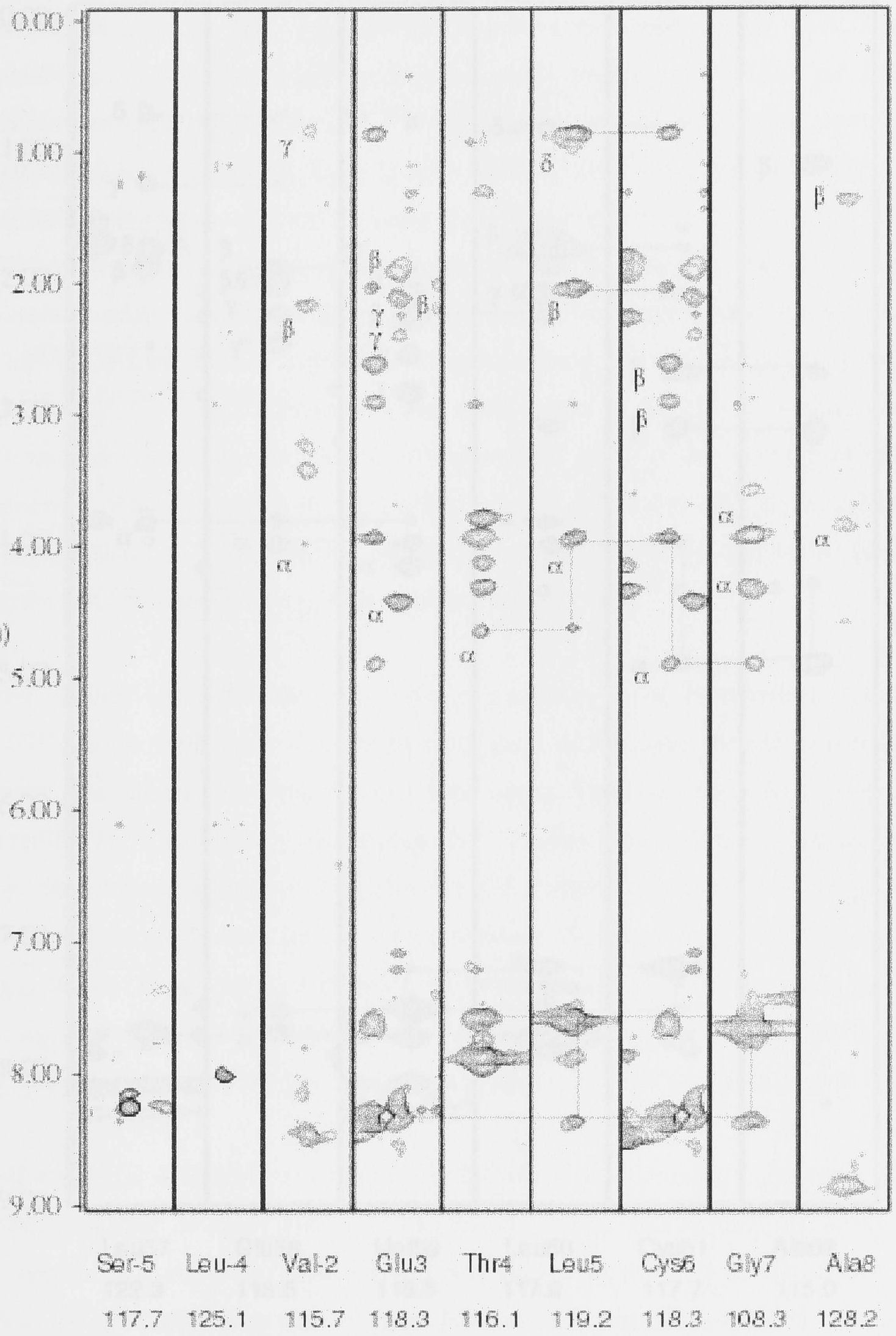

Figure 4.8: Strips from the 3D sensitivity enhanced ^{15}N-NOESY-HSQC spectrum of Long-[Leu^{60}]-IGF-I displaying the sequential contacts observed for the region Ser^{-5} to Ala^{8}.

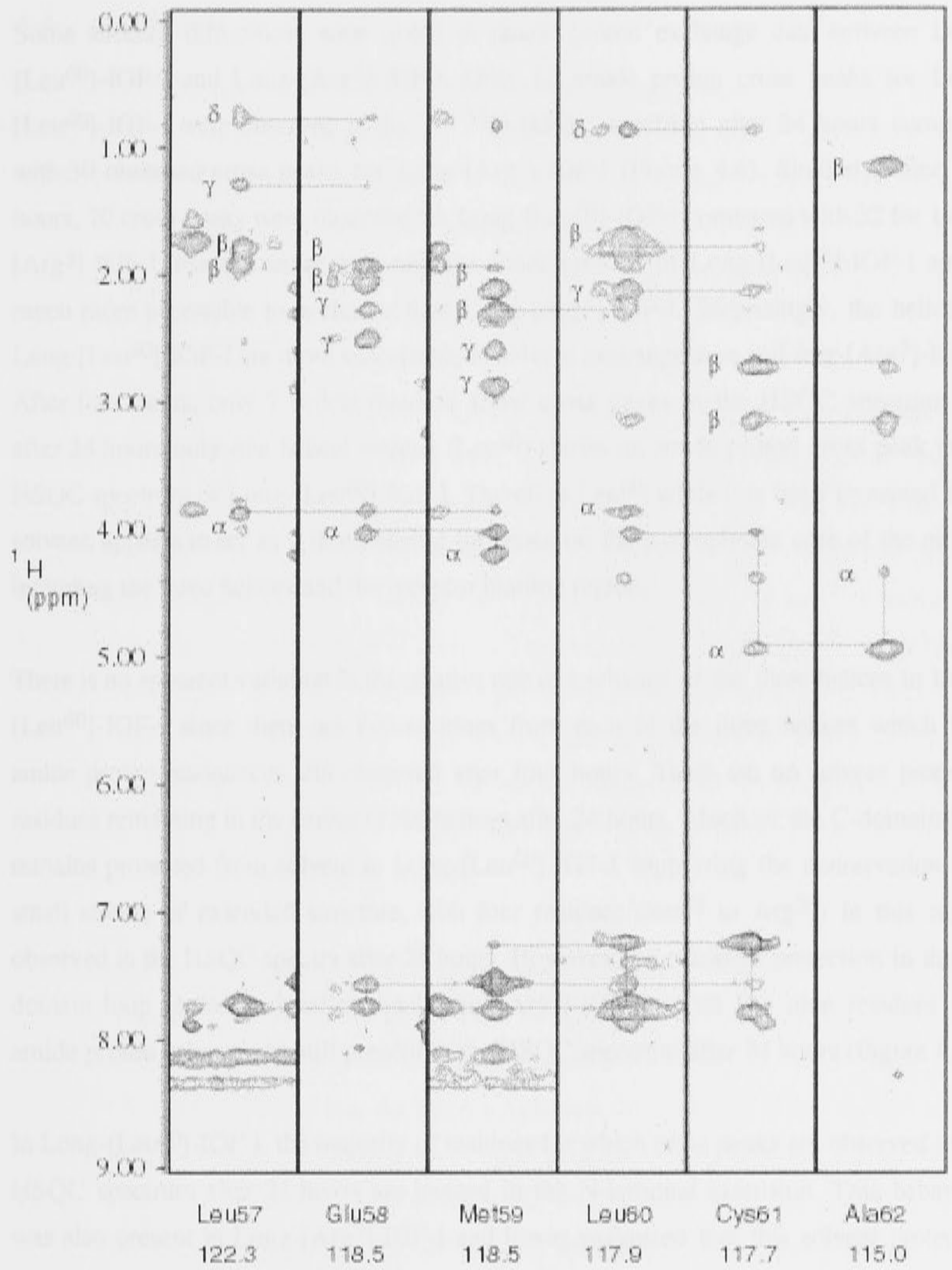

Figure 4.9: Strips from the 3D sensitivity enhanced ^{15}N-NOESY-HSQC spectrum of Long-[Leu^{60}]-IGF-I displaying the sequential and medium-range contacts observed for the region Leu^{57} to Ala^{62}.

Some striking differences were noted in amide proton exchange data between Long-[Leu^{60}]-IGF-I and Long-[Arg^{3}]-IGF-I. Only 12 amide proton cross peaks for Long-[Leu^{60}]-IGF-I were observed in the 2D ^{15}N-HSQC spectrum after 24 hours compared with 30 observed cross peaks for Long-[Arg^{3}]-IGF-I (Figure 4.6). Similarly, after four hours, 20 cross peaks were observed for Long-[Leu^{60}]-IGF-I compared with 32 for Long-[Arg^{3}]-IGF-I. For the entire molecule, the amide protons of Long-[Leu^{60}]-IGF-I appear much more accessible to exchange than Long-[Arg^{3}]-IGF-I. Surprisingly, the helices in Long-[Leu^{60}]-IGF-I are more susceptible to solvent exchange than in Long-[Arg^{3}]-IGF-I. After four hours, only 7 helical residues show cross peaks in the HSQC spectrum and after 24 hours only one helical residue (Leu^{60}) shows an amide proton cross peak in the HSQC spectrum of Long-[Leu^{60}]-IGF-I. Therefore Leu^{60}, while it is itself protected from solvent, appears to act as a destabilising influence on the hydrophobic core of the protein including the three helices and the receptor binding region.

There is no apparent variation in the relative rate of exchange of the three helices in Long-[Leu^{60}]-IGF-I since there are two residues from each of the three helices which have amide proton resonances still observed after four hours. There are no solvent protected residues remaining in the centre of the helices after 24 hours. Much of the C-domain loop remains protected from solvent in Long-[Leu^{60}]-IGF-I supporting the conservation of a small stretch of extended structure, with four residues (Ser^{33} to Arg^{36}) in this region observed in the HSQC spectra after 24 hours. However, the extent of protection in this C-domain loop region is smaller than Long-[Arg^{3}]-IGF-I which has nine residues with amide proton cross peaks still present in the HSQC spectrum after 24 hours (Figure 4.6).

In Long-[Leu^{60}]-IGF-I, the majority of residues for which cross peaks are observed in the HSQC spectrum after 24 hours are located in the N-terminal extension. This behaviour was also present in Long-[Arg^{3}]-IGF-I and it was suggested that this solvent protection may reflect the hydrophobic nature of the extension which arranges itself in transitory structures so as to exclude solvent. However, in Long-[Leu^{60}]-IGF-I, the extent of protection of the N-terminal extension is not as great since the amide proton resonances of only five residues in the N-terminal extension of Long-[Leu^{60}]-IGF-I were present 24 hours after exchange compared with eight N-terminal amide proton resonances in Long-[Arg^{3}]-IGF-I.

4.5 Discussion

A detailed comparison of the NOEs, amide proton exchange rates, J-couplings and H^{α} proton chemical shifts used to determine secondary structure in IGF-I, mini-IGF-I, Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I is presented in this section. Subtle differences in the structural regions or length of secondary structural elements between IGF-I and the analogues may provide partial explanation for differences in biological activity and altered affinity for receptors and binding proteins.

In the first published IGF-I structure (Cooke et al., 1991), the authors identified that helix 1 extended from Ala^8 to Val^{17}, helix 2 from Val^{44} to Phe^{49} and helix 3 from Leu^{54} to Met^{59}. The authors also noted that α-helix 3 is less well defined than the other two, especially at Arg^{55}. A more recent structure of IGF-I (Sato et al., 1992) identified the first helix as encompassing residues Ala^8 to Cys^{18}, one residue longer at the C-terminal end than identified by Cooke et al. (1991). Helix 2 extended from Gly^{42}-Cys^{48}, while helix 3 extended from Leu^{54} to Cys^{61}, with the length of two turns. For mini-IGF-I, helix 1 was found to extend from Ala^8 to Cys^{18}, the same length as identified by Sato et al. (1992), helix 2 extended from Gly^{42} to Phe^{49}, one residue longer at the C-terminus than IGF-I (Sato et al., 1992) while helix 3 extended from Leu^{54} to Cys^{61}, encompassing the same residues as identified by Sato et al. (1992).

The three helices of Long-[Arg^3]-IGF-I encompass residues Ala^8 to Cys^{18}, Ile^{43} to Arg^{50} and Leu^{54} to Tyr^{60} respectively. Helix 1 of Long-[Arg^3]-IGF-I encompasses the same residues as helix 1 of IGF-I (Sato et al., 1992) and mini-IGF-I (De Wolf et al., 1996), helix 2 of Long-[Arg^3]-IGF-I is the same length as helix 2 in mini-IGF-I though shifted by one residue toward the C-terminus. Helix 3 of Long-[Arg^3]-IGF-I is one residue shorter than helix 3 in both IGF-I (Sato et al., 1992) and mini-IGF-I (De Wolf et al., 1996). The three helices of Long-[Leu^{60}]-IGF-I are located in the same positions as the helices of Long-[Arg^3]-IGF-I, though helix 2 of Long-[Leu^{60}]-IGF-I is not as well defined as in Long-[Arg^3]-IGF-I, especially the segment Cys^{47} to Arg^{50}.

In all IGF structural studies, the start and end points of the α-helices were defined by strong d_{NN} (i,i+1), $d_{\alpha N}$(i,i+3), $d_{\alpha N}$(i,i+4) and $d_{\alpha\beta}$(i,i+3) connectivities for IGF-I (Sato et al., 1992), mini-IGF-I (De Wolf et al., 1996), Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I. The differences in the start and end points of helices between IGF-I and the analogues

are small and may be explained by differences in sample and NMR experimental conditions. The magnitude of some NOEs were uncertain in previous homonuclear studies of IGF-I (Cooke et al., 1991; Sato et al., 1992) because of overlap in some regions. In this work, these regions are better defined because of more advanced NMR technology and pulse sequences.

Hydrogen-deuterium exchange experiments are quite sensitive to conformational changes and variations in solvent exposure between proteins. Identification of slowly exchanging backbone amide protons was used to confirm the location of secondary structure in IGF-I (Sato et al., 1992), mini-IGF-I (De Wolf et al., 1996), Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I. Cooke et al. (1991) did not report hydrogen-deuterium exchange rates in their publication of the IGF-I structure. Sato et al. (1992) identified 17 backbone amide protons in IGF-I as slowly exchanging, all of which were in helical regions with the exception of Gly^{22}, which is involved in a hydrogen bond in the β-turn following helix 1. In mini-IGF-I, 18 amide proton resonances were still observed 30 min after sample preparation, all of which were in helical regions with the exception of Gly^{22} (De Wolf et al., 1996). In addition, a further 11 amide proton resonances were noted as undergoing slow exchange. Only one of these residues, Ala^8, was in a helical region. The others were Gly^1, Thr^4, Leu^5, Gly^7, Tyr^{24}, Ser^{51}, Asp^{53}, Leu^{64}, Lys^{65} and Ser^{69}.

Long-[Arg^3]-IGF-I showed a different pattern of solvent exchange with 30 backbone amide proton resonances still present 24 hours after solvent preparation, 12 of these in helical regions. Eight residues in the C-domain of Long-[Arg^3]-IGF-I (Gly^{30}-Ala^{38}) had slowly exchanging amide protons, none of which were slowly exchanging in IGF-I. IGF-II has a smaller stretch of residues in an extended conformation between helix 1 and helix 2 that are protected from solvent, although, compared to Long-[Arg^3]-IGF-I, this region is offset five residues toward the N-terminus at Gly^{25} to Phe^{28} (Torres et al., 1995). An additional 12 residues of Long-[Arg^3]-IGF-I had backbone amide protons still observed after four hours, with seven of these in helical regions. Slowly exchanging amide protons were observed for Leu^5, Gly^7, and Leu^{64} of Long-[Arg^3]-IGF-I, consistent with the amide exchange patterns for mini-IGF-I.

In contrast to Long-[Arg^3]-IGF-I, overall only 12 residues of Long-[Leu^{60}]-IGF-I have amide proton resonances still observed after 24 hours, suggesting that the analogue is less protected from solvent than Long-[Arg^3]-IGF-I, with fewer residues participating in

hydrogen bond interactions. The helical regions of Long-[Leu^{60}]-IGF-I are more exposed to solvent than the same regions of Long-[Arg^{3}]-IGF-I with the resonance of only one backbone amide proton still observed after 24 hours. Residues in the C-domain of Long-[Leu^{60}]-IGF-I are strongly protected from solvent with the backbone amide proton resonances of four residues (Ser^{33}-Arg^{36}) still observed after 24 hours.

$^{3}J_{NH\alpha}$ values were not obtained for IGF-I or mini-IGF-I. This is unfortunate since it would have been useful to know if the series of $^{3}J_{NH\alpha}$ values $\geq$ 7.5 Hz observed for the C-domain segment (Gly^{30} to Ala^{38}) of both Long-[Arg^{3}]-IGF-I and Long-[Leu^{60}]-IGF-I were observed in IGF-I. This region of IGF-I did not display the same solvent protection as Long-[Arg^{3}]-IGF-I and Long-[Leu^{60}]-IGF-I. The analogous solvent protected region in the C-domain of IGF-II did not have $^{3}J_{NH\alpha}$ values $\geq$ 7.5 Hz.

Chemical shift index analyses for the H$^{\alpha}$ protons were not performed by Cooke et al. (1991) or Sato et al. (1992) for IGF-I since the chemical shift index tables of Wishart et al. (1991) were not available at the time. However, tables of ^{1}H chemical shift resonances including H$^{\alpha}$ proton chemical shifts for IGF-I were provided by both research groups (Cooke et al., 1991; Sato et al., 1992), allowing chemical shift index analyses to be performed for the purposes of comparison with Long-[Arg^{3}]-IGF-I and Long-[Leu^{60}]-IGF-I.

Upfield H$^{\alpha}$ chemical shifts are generally observed in the helical regions for IGF-I, mini-IGF-I, Long-[Arg^{3}]-IGF-I and Long-[Leu^{60}]-IGF-I. However, a number of exceptions are noted. The H$^{\alpha}$ resonance of Ala^{8} is shifted downfield in mini-IGF-I relative to the random coil value, whereas this resonance is shifted upfield in the two Long-IGF-I analogues. The H$^{\alpha}$ resonances of Cys^{18}, the C-terminal residue of helix 1, is shifted downfield relative to the random coil value in IGF-I and mini-IGF-I) but not in either of the Long-IGF-I analogues (Cooke et al., 1991; Sato et al., 1992; De Wolf et al., 1996). The H$^{\alpha}$ resonance of Cys^{47} also experiences a downfield shift in IGF-I (Cooke et al., 1991; Sato et al., 1992), mini-IGF-I (De Wolf et al., 1996) and Long-[Arg^{3}]-IGF-I. The Tyr^{60} H$^{\alpha}$ resonance experiences a downfield chemical shift with respect to the random coil value in IGF-I (Sato et al., 1992), mini-IGF-I (De Wolf et al., 1996) and Long-[Arg^{3}]-IGF-I. In addition, the H$^{\alpha}$ resonances of two other Cys residues, Cys^{52} and Cys^{61} experience downfield chemical shifts with respect to the random coil value in IGF-I, mini-IGF-I and Long-[Arg^{3}]-IGF-I (Cys^{61} only). The H$^{\alpha}$ resonance of Leu^{57} (helix 3) in Long-[Arg^{3}]-

IGF-I is shifted downfield with respect to the random coil value. Similarly, the H^{α} resonances of Glu^{46} and Arg^{50} (helix 2) in Long-[Leu^{60}]-IGF-I are shifted downfield, indicating that helix 2 of Long-[Leu^{60}]-IGF-I has experienced some disruption when compared with IGF-I, due to the $Tyr^{60}Leu$ mutation.

Residues in the segment Phe^{23} to Phe^{25} of IGF-I have been identified as important for IGF receptor binding (Cascieri et al., 1988). In Long-[Arg^{3}]-IGF-I, the H^{α} resonances of Phe^{23} to Asn^{26} all experience a downfield shift with respect to their random coil values. However, in Long-[Leu^{60}]-IGF-I, the H^{α} resonances are shifted upfield with respect to the random coil values for the segment Tyr^{24} to Lys^{27}. These chemical shift differences are presented in Table 4.2. Long-[Leu^{60}]-IGF-I binds less well to the type 1 IGF receptor than IGF-I (Milner et al., 1999). The $Tyr^{60}Leu$ mutation is the major contributing factor to decreased binding of Long-[Leu^{60}]-IGF-I to the type 1 IGF receptor. However, it is difficult to distinguish the ring current effect from a structural effect.

Table 4.2. H^{α} chemical shift deviations from random coil H^{α} shifts

Residue	**Long-[Arg^{3}]-IGF-I**	**Long-[Leu^{60}]-IGF-I**
Phe^{23}	+0.52	+0.28
Tyr^{24}	+0.22	-0.11
Phe^{25}	+0.15	-0.16
Asn^{26}	+0.25	-0.26
Lys^{27}	0.00	-0.20

The stability and biological activity of globular proteins are dictated by the close packing of amino acids. Disruption of the packed core of a protein, even in the form of quite subtle mutations, can result in dramatic effects (reviewed by Shortle (1989), Fersht & Serrano (1993) and Sturtevant (1994)). Rearrangements of the main chain of proteins can allow accommodation of substituted residues. Major disruptions in the packing interactions occur when polar charged residues are introduced into or removed from the protein core (Ladbury et al., 1995). In this study, neither IGF-I substituted residues, Glu^{3} or Tyr^{60}, are buried in the protein core, although Tyr^{60} is positioned at the C-terminal end of α-helix 3. The $Glu^{3}Arg$ substitution in Long-[Arg^{3}]-IGF-I, despite involving a change in charge,

does not result in a large structural reorientation of the IGF-I domain. Rather, the substitution results in local conformational changes at the N-terminus (see section 5.4.5). From the NMR data presented herein, the Tyr^{60}Leu substitution in Long-[Leu^{60}]-IGF-I does result in some chemical shift changes in nearby secondary structure elements but this does not disrupt the protein core. This was also observed for a mutant of thioredoxin, Leu^{78}Lys, in which structural changes caused by the mutation were localized within 12 Å of the altered sidechain (De Lorimier et al., 1996). However, this mispacking of the protein core was reported to affect the local dynamics and stability throughout the protein. The mobility and dynamics of the Long-IGF analogues were investigated by NMR spectroscopy and the results are discussed in Chapter 6.

The analysis of NOE contacts, J-coupling information, amide proton exchange rates and chemical shift index data confirms that the location of the secondary structure elements of the Long-IGF analogues is essentially unchanged from the parent IGF-I molecule.

CHAPTER 5

Three-Dimensional Solution Structure of Long-[Arg^3]-IGF-I

5.1 Introduction

Three-dimensional solution structure calculations were performed for the Long-[Arg^3]-IGF-I analogue but not for the Long-[Leu^{60}]-IGF-I analogue for the following reasons. Firstly, the binding studies between Long-[Leu^{60}]-IGF-I and IGFBP-2 (described in chapter 7) were considered to be a priority for this analogue. Secondly, the resonance assignments and secondary structure of Long-[Leu^{60}]-IGF-I were not sufficiently different from Long-[Arg^3]-IGF-I (section 4.4.2) to indicate that the three-dimensional structure would be substantially changed. Finally, no unlabelled Long-[Leu^{60}]-IGF-I was available, meaning that unambiguous identification of long-range contacts between sidechain protons could not be completed. Thus, all subsequent sections of this chapter refer only to the structure of Long-[Arg^3]-IGF-I.

The major interest in the structure of Long-[Arg^3]-IGF-I resides in the effects that the thirteen amino acid extension and the Glu^{3}Arg mutation have on the three-dimensional structure of the IGF-I domain. As discussed in section 1.8, Long-[Arg^3]-IGF-I is among the most potent of the IGF-I analogues with N-terminal modifications. Enhanced biological activity is observed despite a lower affinity for the type 1 IGF receptor compared to wild-type IGF-I. The increased potency has been attributed to reduced binding to IGFBPs, which increases the concentration of free Long-[Arg^3]-IGF-I available to the receptor. A three-dimensional model of Long-[Arg^3]-IGF-I (presented in this chapter) enables a partial rationalisation of a wealth of experimental data collected over the past decade with regard to the binding of IGF-I and its analogues to IGFBPs, the IGF receptors and the insulin receptor.

5.2 Materials and Hardware

5.2.1 Software

Three-dimensional NOESY cross peaks were integrated using the program SPSCAN (R.W.Glaser and K.Wüthrich, http://www.biol.ethz.ch/wuthrich/software/spscan/). CALIBA (part of the DYANA suite of programs) was used to convert NOESY cross peak volumes and disulfide bonds into upper and lower distance bounds. Automatic calibration was performed by the macro CALIBA. ϕ, ψ and χ_1 angle restraints and H$^\beta$ assignments

were created from experimentally derived $^3J_{HNH\alpha}$ and $^3J_{HNH\beta}$ values using the program HABAS (Güntert et al., 1989). Structures were generated using the program DYANA 1.5 (Güntert et al., 1997) which employs simulated annealing combined with molecular dynamics in torsion angle space. DYANA calculates structures by adjusting torsion angles to fit conformational restraints derived from experimental data while all covalent structure parameters (bond lengths, bond angles, chiralities and planarities) are kept at their optimum values. The quality of the structures were analysed using the programs AQUA and PROCHECK-NMR (Laskowski et al., 1996).

5.2.2 Sedimentation Equilibrium Apparatus

A Beckman Optima XL-A analytical centrifuge (Beckman Instruments, Fullerton, CA) located at the John Curtin School of Medical Research, Australian National University, was used for the sedimentation equilibrium experiments. 12 mm path length cells with aluminium filled double sector centrepieces were used.

5.3 Methods

5.3.1 Sedimentation Equilibrium

The solution sector of the analytical centrifuge contained 100 µl of a 1 mM sample of Long-[Arg^3]-IGF-I dissolved in H_2O while the solvent sector contained 110 µl of H_2O as a reference. Experiments were performed at 25 ^oC. The equilibrium profile of absorbance of the protein, A, at 280 nm as a function of the radial distance (in cm) from the axis of rotation at the centre of the rotor, r, was recorded. The attainment of sedimentation equilibrium was checked by collecting and superimposing scans at two hourly intervals. Scans at 360 nm were collected and subtracted from the equilibrium scans to correct for anomalies arising from the cell windows.

5.3.2 NMR Experimental

The HNHA experiment (Vuister and Bax, 1993) was used to determine $^3J_{NH\alpha}$ coupling constants and was described in section 4.3.1. The HNHB experiment (Archer et al., 1991) was used to correlate amide proton (HN) and ^{15}N resonances with those of intraresidual sidechain $H\beta$ proton signals. Measurement of the cross-peak intensity provides qualitative information on the size of $^3J_{NH\beta}$. In this experiment, H^β cross peaks have negative

intensities while H^N diagonal cross peaks have positive intensities. This yields information about the sidechain torsion angle χ_1 and aids in the stereospecific assignment of non equivalent methylene protons. The pulse scheme is presented in Appendix 1. A 3D ^{15}N-NOESY-HSQC spectrum with mixing time 80 ms and a 2D NOESY spectrum with mixing time 120 ms were used for assigning upper and lower distance restraints.

5.3.3 Experimental Restraints

Angular restraints, derived from $^3J_{HNH\alpha}$ via the Karplus relation (equation 5.1), were used to restrict the sampling of conformational space by the protein backbone during structure refinement. The magnitude of the ^{3}JHNHα values is related to the backbone dihedral angle of a protein, θ (shown in Figure 1.5).

$$^3J = A\cos^2\theta + B\cos\theta + C \qquad (5.1)$$

The constants A, B, and C depend on the particular nuclei that form three adjoining covalent bonds. Accurate parameterisation of the Karplus equation is necessary in order to relate measured ^{3}J values to dihedral angle ranges. The most accurate values for the constants A, B, and C have been derived by correlation of observed ^{3}J values with the corresponding dihedral angles measured in protein structures determined by X-ray crystallography or NMR spectroscopy (Pardi et al., 1984; Ludvigsen, 1991). Different conformations can yield the same value of $^3J_{HNH\alpha}$, although some of the possible orientations may be sterically unfavourable. Dihedral angle constraints enforced during structural calculations must include all of the valid solutions to the Karplus curve (Cavanagh et al., 1996). $^3J_{HNH\alpha}$ values observed in the range 6-8 Hz generally are not used as protein structural restraints because they may reflect motional averaging of multiple conformations.

Two methods were used for restraining phi (ϕ) and psi (ψ) angles. Firstly, for $^3J_{NH\alpha} \leq$ 5.0 Hz, ϕ was restrained to -60 ± 30°, for values 5.0 Hz $\leq$ $^3J_{NH\alpha} \leq$ 5.5 Hz, ϕ was restrained to -60 ± 40° and for values of $^3J_{NH\alpha} \geq$ 7.5 Hz, ϕ was restrained to -120 ± 30°. ϕ was not restrained for $^3J_{NH\alpha}$ values between 5.5 and 7.5 Hz. The second method employed HABAS which determines the allowed conformation for the ϕ, ψ, and χ_1 dihedral angles using a grid search on the basis of local distance restraints and scalar coupling constants. The upper and lower dihedral angle restraints generated by

both methods were inspected to ensure that they were self consistent.

χ_1 angle restraints and H^β stereospecific assignments were made using the program HABAS via estimates of the $^3J_{NH\beta}$ values obtained from the relative HN-H^β cross peak intensities in a 3D HNHB spectrum. $^3J_{NH\beta}$ values were treated qualitatively and given the limits of ± 2.0 Hz. If one of the HN-H^β cross peaks is stronger than the other in the HNHB experiment, this reflects the presence of a large and small coupling constant. A difference in strength of the cross peaks was defined as one cross peak being three times as intense as the other cross peak. The larger cross peak was assigned a $^3J_{NH\beta}$ value of 5.0 Hz while the smaller coupling constant was assigned a $^3J_{NH\beta}$ value of 1.5 Hz. If only one HN-H^β cross peak was observed, $^3J_{NH\beta}$ was assigned a value of 5 Hz. If weak or vanishing HN-H^β correlations were observed for both β-protons or both cross peaks were equally intense, no input was given for these results. These $^3J_{NH\beta}$ values were assigned in accordance with the Karplus relationship (equation 5.1).

Slowly exchanging backbone amide protons were identified as described in section 4.3.3. Hydrogen bond restraints were included for slowly exchanging NH groups, for which a single hydrogen bond acceptor was identified in preliminary structure calculations, and was consistent with the secondary structure determined previously. Lower distance restraints between the hydrogen bond donor and acceptor and the hydrogen bond donor antecedent and the acceptor of 1.5 Å and 2.3 Å respectively were added. Upper distance restraints for the same bonds were 2.3 Å and 3.3 Å respectively.

Upper distance bounds were calculated from the integration of three-dimensional NOESY cross peaks using the program SPSCAN as described earlier (section 5.2.1).

5.3.4 Structure Calculations

An iterative process of calculation was employed where successive rounds of structure calculations were used to make further NOE assignments. During preliminary rounds of structure calculations, typically 50-100 structures were generated using 10000 simulated annealing steps followed by 1000 energy minimisation steps, with 15 or 25 structures being retained for analysis of hydrogen bonds and long range NOEs. The 15 sets of coordinates with the lowest target functions were selected from 100 coordinate sets in the final cycle of calculations.

5.3.5 Evaluation of the Quality of the Structure

AQUA is a suite of programs for analysing the quality of biomolecular structures determined by NMR spectroscopy. PROCHECK-NMR is a suite of programs that produce PostScript plots of the restraint violations computed by AQUA together with various analyses of the stereochemical quality of the models in the NMR ensemble. The PDB file (Bernstein et al., 1977) with the coordinates of Long-[Arg^3]-IGF-I, a restraint file listing NOEs, hydrogen bonds, disulphides and torsion angle restraints were input for the two programs AQUA and PROCHECK. The angular order parameters give an indication of how well defined the backbone dihedral angles are. An angular order parameter close to one represents a well-defined backbone dihedral angle, while low angular order parameters indicate poorly defined backbone dihedral angles. Angular order parameters for Long-[Arg^3]-IGF-I were determined using MOLMOL (Koradi et al., 1996).

5.4 Results

5.4.1 Sedimentation Equilibrium

Sedimentation equilibrium was reached after seven hours at a speed of 3000 rpm. The resulting absorbance (A) vs radial distance (r) in cm from the axis of rotation data were analysed using the program XLAEQ (Beckman Instruments, Fullerton, CA), employing the sedimentation equilibrium equation:

$$M=2RT/(1-v\rho)\omega^2 . d(\ln A)/(dr^2) \qquad (5.2)$$

where M is the molecular weight, R is the gas constant, T is the absolute temperature, ω is the angular rotation in rad/sec, v is the partial specific volume of the protein, 0.73 ml/g in this case, and d is the density of the buffer. For a non-associating, monodisperse system, the plot of ln A vs r^2 will be linear. The data could not be fitted to a single straight line using equation (5.2), the plot of the data was curved upwards, indicating a heterogenous protein (Figure 5.1). The apparent heterogeneity in molecular weight of the protein was examined to determine if this could be attributed to association of the protein. If the system is self-associating or involved in 'heterologous' association (i.e. complex formation) the ln A vs r^2 plot can be used to measure the stoichiometry and strength of the

interaction. Assuming a mixture of monomer and dimer, using the program MULTMX1 (Beckman Instruments, Fullerton, CA), a good fit was obtained for a mixture of 80% monomer (mol.wt. 9000) and 20% dimer (mol. wt. 18000).

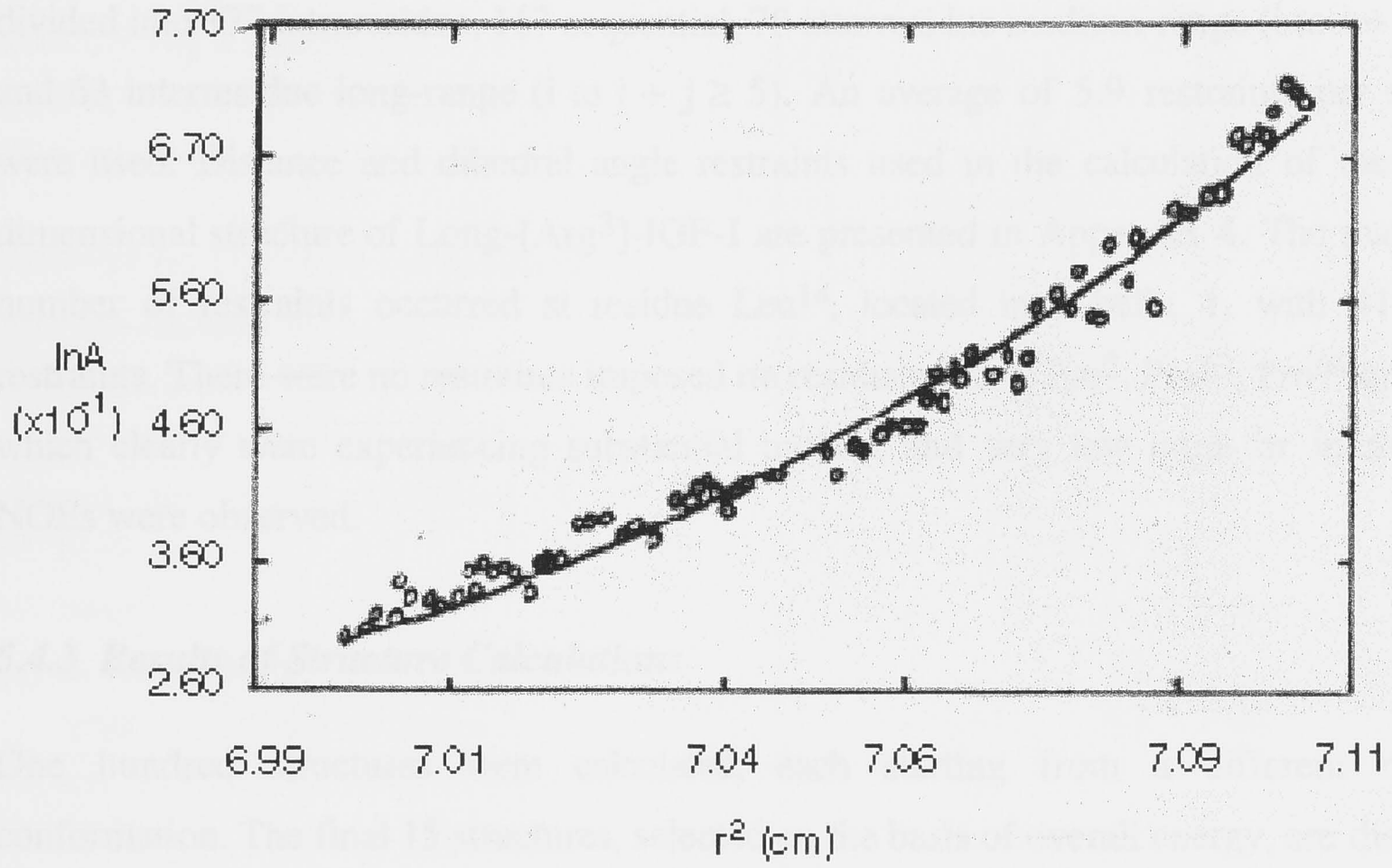

Figure 5.1: Plot of lnA vs r^2 for a 1 mM solution of Long-[Arg^3]-IGF-I.

5.4.2 Experimental Restraints

The input restraints were obtained from 3D ^{1}H-^{15}N NOESY-HSQC and 2D ^{1}H NOESY experiments. Initially, structures were calculated using only 389 unambiguous experimentally derived distance restraints, 90 restraints derived from coupling constants and 12 restraints to maintain the three disulfide bridges. Subsequently, 24 upper bound and 24 lower bound hydrogen bond restraints were added to maintain 12 hydrogen bonds that were deduced from 12 very slowly exchanging amide protons. These hydrogen bonds were primarily within the three α-helices with one restraint added for a hydrogen bond in the type 1 β-turn following α-helix 1 (Gly^{19}-Gly^{22}). An additional 107 NOEs were unambiguously assigned and added as restraints after inspection of the preliminary

structures. The stereospecific assignments of nine methylene β protons were deduced by inspection of the relative crosspeak intensities in the 3D HNHB, ^{1}H-^{15}N NOESY-HSQC and ^{1}H-^{15}N TOCSY-HSQC spectra. There was no evidence f intermonomer NOEs and so no restraints were expected to be due to dimerization.

The final structures were calculated on the basis of 524 distance restraints, 90 dihedral angle restraints and 48 hydrogen bond restraints. Interproton distance restraints could be divided into 177 intraresidue, 167 sequential, 78 interresidue medium-range (i to i + j ≤ 4) and 63 interresidue long-range (i to i + j ≥ 5). An average of 5.9 restraints per residue were used. Distance and dihedral angle restraints used in the calculation of the three-dimensional structure of Long-[Arg^{3}]-IGF-I are presented in Appendix 4. The maximum number of restraints occurred at residue Leu^{14}, located in α-helix 1, with 41 NOE restraints. There were no restraints imposed on residues Pro^{-8}, Pro^{2}, Pro^{63}, Pro^{66} or Ala^{70}, which clearly were experiencing substantial motion, and very few intra- or interresidue NOEs were observed.

5.4.3 Results of Structure Calculations

One hundred structures were calculated, each starting from a different random conformation. The final 15 structures, selected on the basis of overall energy, are displayed in Figure 5.2. Each structure satisfies the experimental restraints with no upper-bound violations greater than 0.6 Å, no lower-bound violations greater than 0.1 Å and no dihedral angle violations greater than 5°. The backbone conformations are well defined except for the first nine residues of the N-terminal extension, the final eight residues of the C-terminus and residues Asn^{26} to Pro^{39} in the C-domain loop between the first and second helices. The overall mean global backbone rmsd value for the 15 best structures is 3.5 ± 0.9 Å (excluding the N-terminal nine residues and the C-terminal eight residues) and 0.82 ± 0.14 Å for the helical regions. The heavy atom rmsd values for the equivalent regions are 4.4 ± 0.9 Å and 1.38 ± 0.33 Å respectively. The backbone dihedral angles (ϕ and ψ) are well defined, with angular order parameters (S ≥ 0.8) for residues comprising the hydrophobic core of the protein (Figure 5.3), i.e. Ala^{8} to Phe^{23}, Gln^{40} to Cys^{52} and Leu^{54} to Ala^{62}, as well as Ser^{-6} to Ser^{-5}, Phe^{-3}, Leu^{4}, Cys^{6}, Lys^{27}, Thr^{29}, Gly^{32} and Ser^{35}. A survey of the structural statistics and residual violations of the experimental restraints is displayed in Table 5.1.

Table 5.1. Structural statistics and root mean square deviations for the 15 lowest energy solutions of human Long-[Arg^3]-IGF-I

Structural Restraints	
Distance restraints	
total	485
intraresidue	177
sequential	167
medium-range $\|i-j\| \leq 5$	78
long-range $\|i-j\| \geq 5$	63
Dihedral angle restraints	90
Hydrogen bond restraints	48
Disulfide bond restraints	18
Statistics for Structure Calculations	
Average restraint violations	
NOE violation (Å)	0.17 ± 0.22
dihedral angle violation (°)	1.63 ± 0.52
Atomic rmsd values, all residues (Å)	
backbone	5.3 ± 1.6
range	3.16-9.66
all heavy atoms	6.0 ± 1.4
range	3.67-9.99
Atomic rmsd values, residues ñ3 to 62 (Å)	
backbone	3.5 ± 0.9
all heavy atoms	4.4 ± 0.9
Atomic rmsd values, α-helices (Å)	
backbone	0.82 ± 0.28
all heavy atoms	1.71 ± 0.34
Target Function ($Å^2$)	55.16 ± 1.93

5.4.4 Quality of the Structure

A distribution of the NOE restraints as a function of residue number, the rmsd value as a function of residue number and the distribution of angular (ϕ and ψ) order parameters for Long-[Arg^3]-IGF-I are shown in Figure 5.3. The Ramachandran plot of the average minimised structure (Figure 5.4) shows that 43% of the meaningful residues fall into the most favoured regions and 41% of residues are in other allowed regions. The average structure was calculated using molmol. The remaining 16% of residues, all of which are in the disordered terminal and loop regions of the protein, fall in the so-called generously allowed region of the Ramachandran plot. No angular values occur in the disallowed

5.4.5 Description of the Solution Structure

The solution structures of Long-[Arg^3]-IGF-I display good convergence to a single fold in the hydrophobic core region containing the three α-helices (Figure 5.2). A file containing the coordinates of the ensemble of the 15 structures with the lowest target functions have been submitted to the Protein Data Bank. The coordinates entry for Long-[Arg^3]-IGF-I has the PDB ID code 3lri and the NMR restraints entry has PDB ID code 3lrimr. The structured regions in the IGF-I domain of Long-[Arg^3]-IGF-I are retained from the parent protein, in which helix 2 and helix 3 run approximately anti-parallel to each other and roughly perpendicular to helix 1, the three helices enclosing a hydrophobic core (Figure 5.6). This arrangement arises because several long-range NOE interactions were observed between the three α-helices (49% of the long-range NOEs connected pairs of helices) or between residues identified as important for receptor binding, at the C-terminal end of the B-domain.

Overall, the long-range NOEs observed between elements of recognisable secondary structure are almost identical to those observed by Sato et al. (1992) in IGF-I. However, some long-range NOEs were observed in Long-[Arg^3]-IGF-I that were not reported for IGF-I which contribute to subtle changes in the orientation of helix 3 with respect to helix 1. These long-range contacts include NOEs between Val^{11} and Val^{44} and between Leu^{14} and Leu^{57} which were also detected in mini-IGF-I (De Wolf et al., 1996). A related long-range NOE is observed between Leu^{14} and Ile^{43} in Long-[Arg^3]-IGF-I whereas the equivalent NOE for IGF-I is between Ala^{13} and Ile^{43} and in mini-IGF-I this NOE is between Leu^{14} and Val^{44}. In Long-[Arg^3]-IGF-I, an NOE is observed between Asn^{26} and Tyr^{60} whereas NOEs between Tyr^{24} and Tyr^{60} and between Gly^{22} and Tyr^{60} were observed in IGF-I (Sato et al., 1992) but not in Long-[Arg^3]-IGF-I. The retention of the hydrophobic core of the IGF-I domain of Long-[Arg^3]-IGF-I indicates that modifications at the N-terminus have a minimal impact on the overall protein conformation except for a slight reorientation of the three helices and a realignment of residues near the N-terminus. The sidechains of all polar and charged residues of helices 1, 2 and 3 face the solution, i.e. Glu^9, Asp^{12}, Gln^{15}, Asp^{45}, Glu^{46}, Arg^{50}, Arg^{55}, Arg^{56} and Glu^{58} although Glu^{46} and Arg^{55} do not point directly away from the hydrophobic core. Most of the helical residues in Long-[Arg^3]-IGF-I prefer helical locations in proteins, with the exception of Tyr^{60}, as Tyr is a poor helix former (Richardson & Richardson, 1988). However, the aromatic ring of Tyr^{60} faces away from the hydrophobic core.

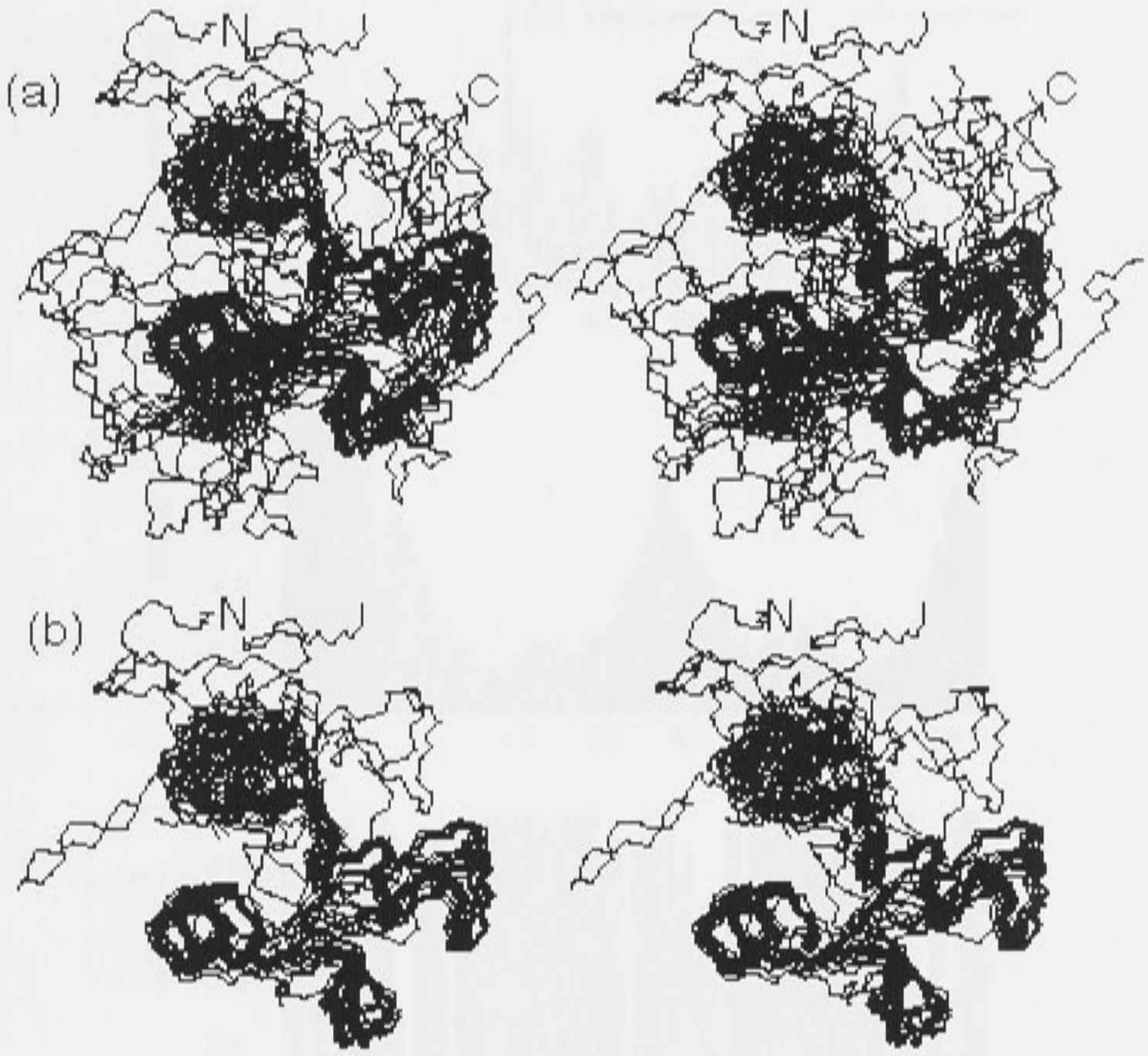

Figure 5.2: Stereo superposition of the backbone heavy-atoms N, C^{α}, C' and O from the final ensemble of 15 structures for Long-[Arg^{3}]-IGF-I. View a) shows all the residues. In view b), the residues in the C-domain loop (Asn^{26}-Gly^{42}) and the C-terminus (Cys^{61}-Ala^{70}) are deleted for clarity.

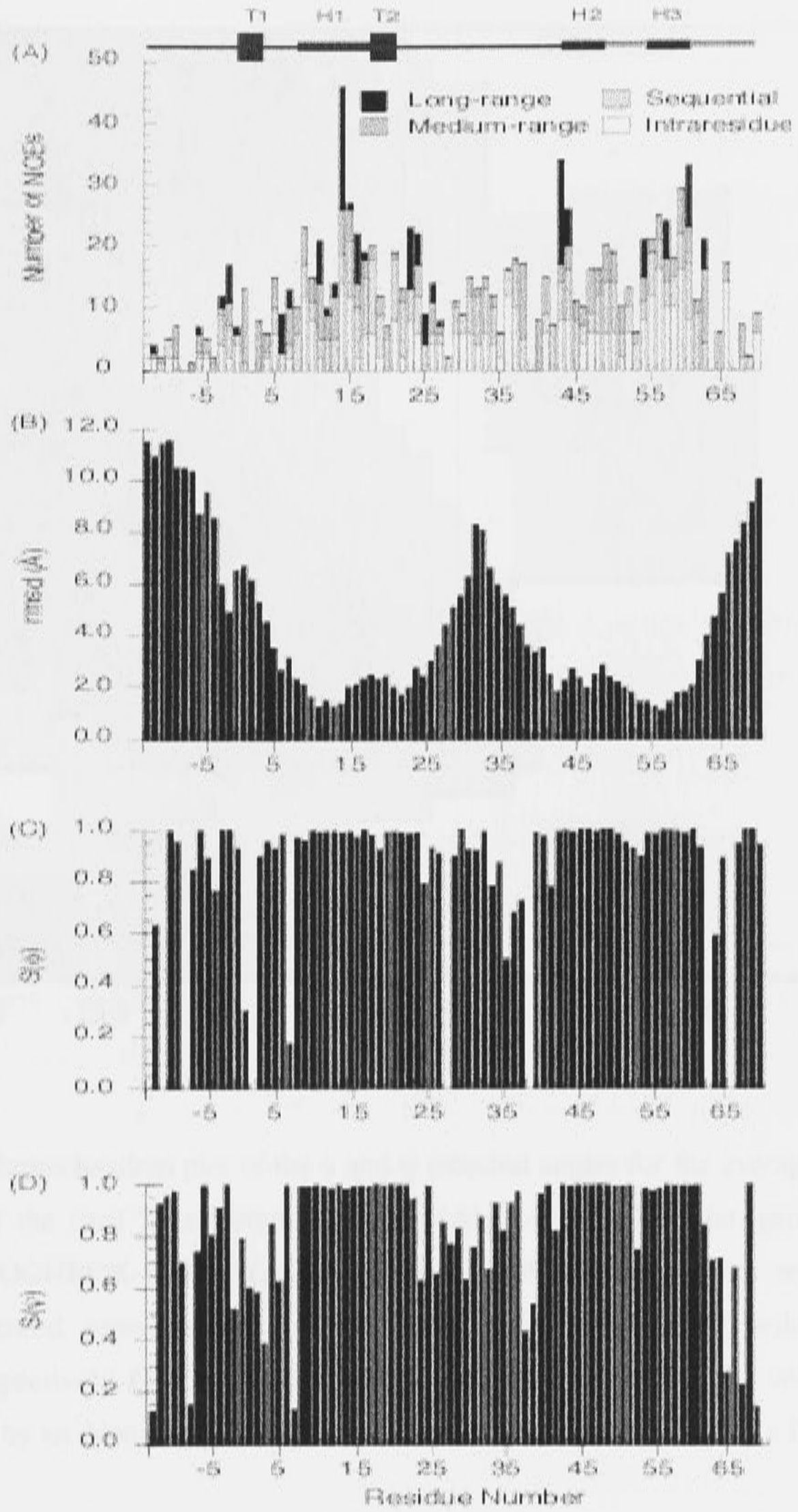

Figure 5.3: a) Distribution of NOE restraints for Long-[Arg^3]-IGF-I. The height of each bar denotes the number of NOEs. b) The rmsd value per residue of the heavy atom backbone coordinates for the 15 best structures of Long-[Arg^3]-IGF-I relative to the average structure. c) Distribution of ϕ angular order parameters, $S(\phi)$, for Long-[Arg^3]-IGF-I. d) Distribution of ψ angular order parameters, $S(\psi)$, for Long-[Arg^3]-IGF-I.

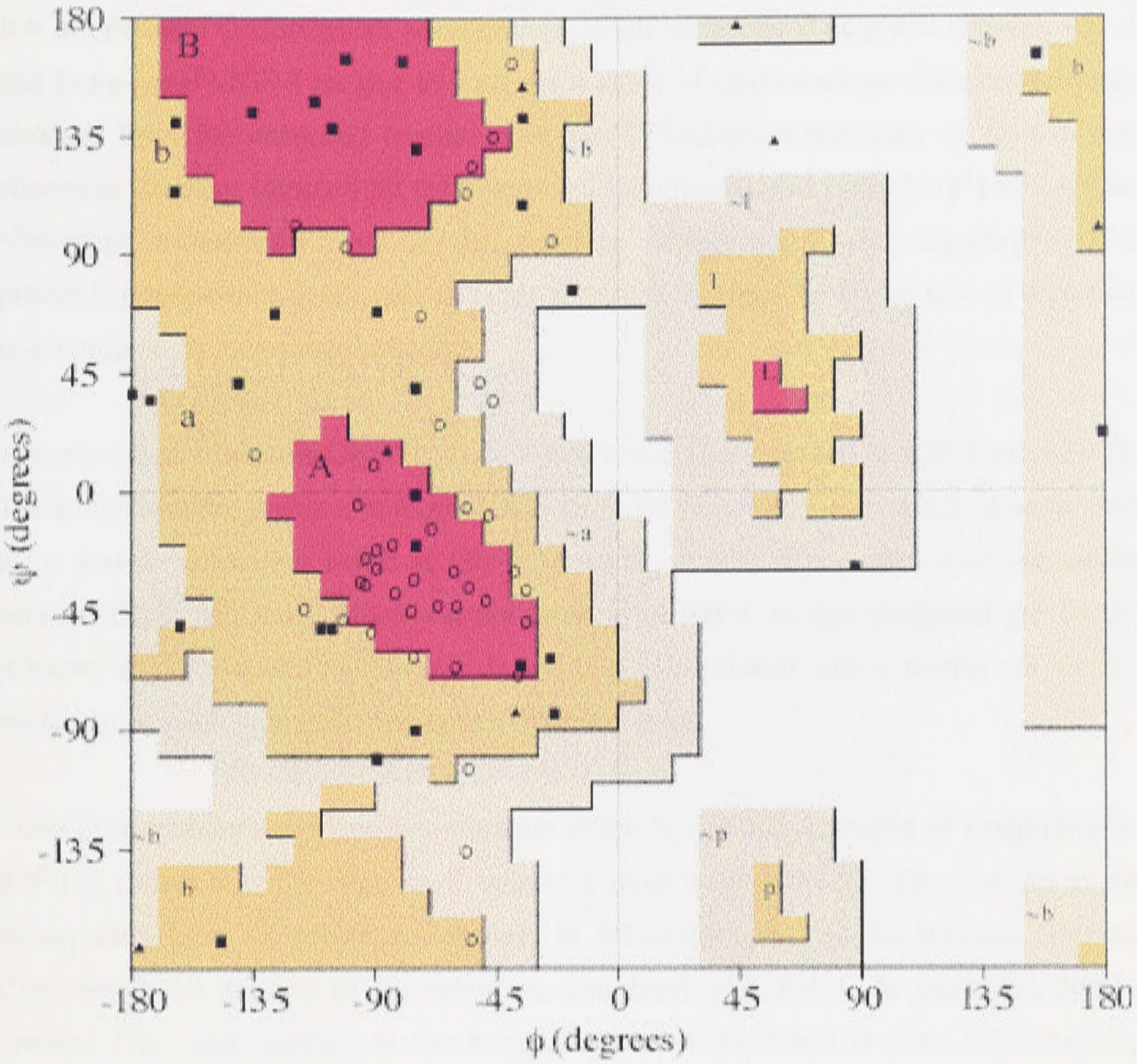

Figure 5.4: Ramachandran plot of the ϕ and ψ dihedral angles for the average structure of the family of the final 15 structures of Long-[Arg^3]-IGF-I. The plot, generated by the program PROCHECK-NMR (Laskowski et al., 1996), displays the most favoured, additional allowed, generously allowed and disallowed regions as red, yellow, light pink and white respectively. Glycine residues are shown as triangles. Residues with rmsd < 4.0 are indicated by an open circle. Residues with rmsd > 4.0 are indicated by a filled square.

It is not possible to distinguish whether these small variations in structure between IGF-I and Long-[Arg^3]-IGF-I are due to a greater number of medium-range distance restraints obtained from the enhanced resolution of the ^{15}N-edited experiments or from a real change in structure imposed by the N-terminal modifications in Long-[Arg^3]-IGF-I. The ^{15}N-edited experiments used in the structure determination of Long-[Arg^3]-IGF-I primarily provide short- and medium-range restraints whereas the global fold of a protein is determined by long-range restraints.

There is a β-turn between Gly^{19} and Gly^{22} which was also observed in IGF-I and IGF-II. The ϕ and ψ angles in this turn most closely fit those of a type 1 turn which is supported by a slowly exchanging amide at Gly^{22}. A small stretch of extended structure exists between Gly^{30} and Ala^{38} which was not detected in IGF-I as was discussed previously (Chapter 4) but was observed in IGF-II and is also consistent with a number of slowly exchanging amide protons in this stretch of amino acids.

There is no well defined secondary structure in the N-terminal extension of Long-[Arg^3]-IGF-I as reflected by the large rmsd values for most residues in the extension. As might be expected, there are substantial changes in the conformation of the residues Gly^1 to Cys^6, which are adjacent to the extension, compared with IGF-I. In particular, NOEs between Gly^1 and residues in the loop between helices 2 and 3 (i.e. Arg^{50}-Asp^{53}), observed in IGF-I, are not present in the spectra of Long-[Arg^3]-IGF-I. Instead, several long range NOEs are observed from Phe^{-3} and Val^{-2} to residues immediately prior to helix 1 (Cys^6 and Gly^7) which places the N-terminal extension near the beginning of helix 1. In Long-[Arg^3]-IGF-I, the residues Asn^{-1}, Gly^1, Pro^2 and Arg^3 form a turn-like structure. Analysis of the ϕ and ψ angles showed that this turn did not belong to any of the published β-turns. The paucity of medium- and long-range NOEs for residues Gly^1 to Thr^4 is reflected in an increase in the rmsd values for these residues (Figure 5.3). Most significant is the dramatic change in orientation of the sidechain of the residue at position 3. In IGF-I, Glu^3 is located in a region bounded by helix 1, helix 3 and the link between helix 2 and helix 3 (Figure 5.5). In Long-[Arg^3]-IGF-I, the sidechain of Arg^3 is located in an exposed position on the surface of the protein (Figure 5.5).

The solution structure of Long-[Arg^3]-IGF-I suggests that the N-terminal extension may fold back over the hydrophobic core of the protein, however, no long-range NOEs are identified between the N-terminal extension and this helical region of the IGF-I domain.

As discussed previously (Chapter 4), there are several slowly exchanging amide protons in the N-terminal extension despite the lack of NOEs, suggesting that these residues arrange themselves in such a way as to exclude solvent. These structures must, however, be transitory because of a lack of medium- and long-range NOEs. The equivalent region of porcine growth hormone, from which the stretch of residues in the N-terminal extension is derived, forms an α-helix (Abdel-Meguid et al., 1987). Thus, the N-terminal extension may retain some of the structure of the corresponding segment of growth hormone and fold into a molten helix in Long-[Arg^3]-IGF-I. The amide protons of Phe^{49} and Arg^{50} near the C-terminus of helix 2 and Arg^{55} and Arg^{56} at the N-terminus of helix 3 are strongly protected from exchange with the solvent, unlike the equivalent residues in native IGF-I (Sato et al., 1992), further supporting the notion that the N-terminal extension does have some order and may form transitory hydrophobic contacts between this region and the N-terminal extension.

5.5 Discussion

5.5.1 Comparison of Long-[Arg^3]-IGF-I with IGF-I and mini-IGF-I

The number of restraints used in the refinement of the structure of Long-[Arg^3]-IGF-I is greater than the number of restraints used for the original IGF-I structures (Cooke et al., 1991; Sato et al., 1993) and is comparable to the number of restraints used for subsequent work with IGF-II (Torres et al., 1995) and mini-IGF-I (De Wolf et al., 1996). The paucity of NOEs in the N-terminal extension contributed to an inflated rmsd value for the entire protein. This is supported by the obvious flexibility in the N-terminal extension (discussed in chapter 6). The data, however, were sufficient to define the global fold, including the position of the three helices and the less mobile part of the N-terminal extension. There was no evidence of intermonomer NOEs.

The fold of the IGF-I domain of Long-[Arg^3]-IGF-I does not substantially differ from IGF-I. However, there are some variations worth noting. A comparison of Long-[Arg^3]-IGF-I with the IGF-I structure of Cooke et. al. (1991) is displayed in Figure 5.5. Immediately obvious is the variation in the orientation of the axes of helix 1 and helix 3. The angles between helices were calculated using MOLMOL (Koradi et al., 1996). The angle between helices 1 and 3 in Long-[Arg^3]-IGF-I is 63.2^o while in IGF-I the same interhelical angle is 87.4^o. The angles between helices 1 and 2 and between helices 2 and 3

are similar in the two proteins, i.e. the angle between helices 1 and 2 is 109.6° in Long-[Arg^3]-IGF-I and 106.0° in IGF-I and the angle between helices 2 and 3 is 160.3° in Long-[Arg^3]-IGF-I and 166.4° in IGF-I.

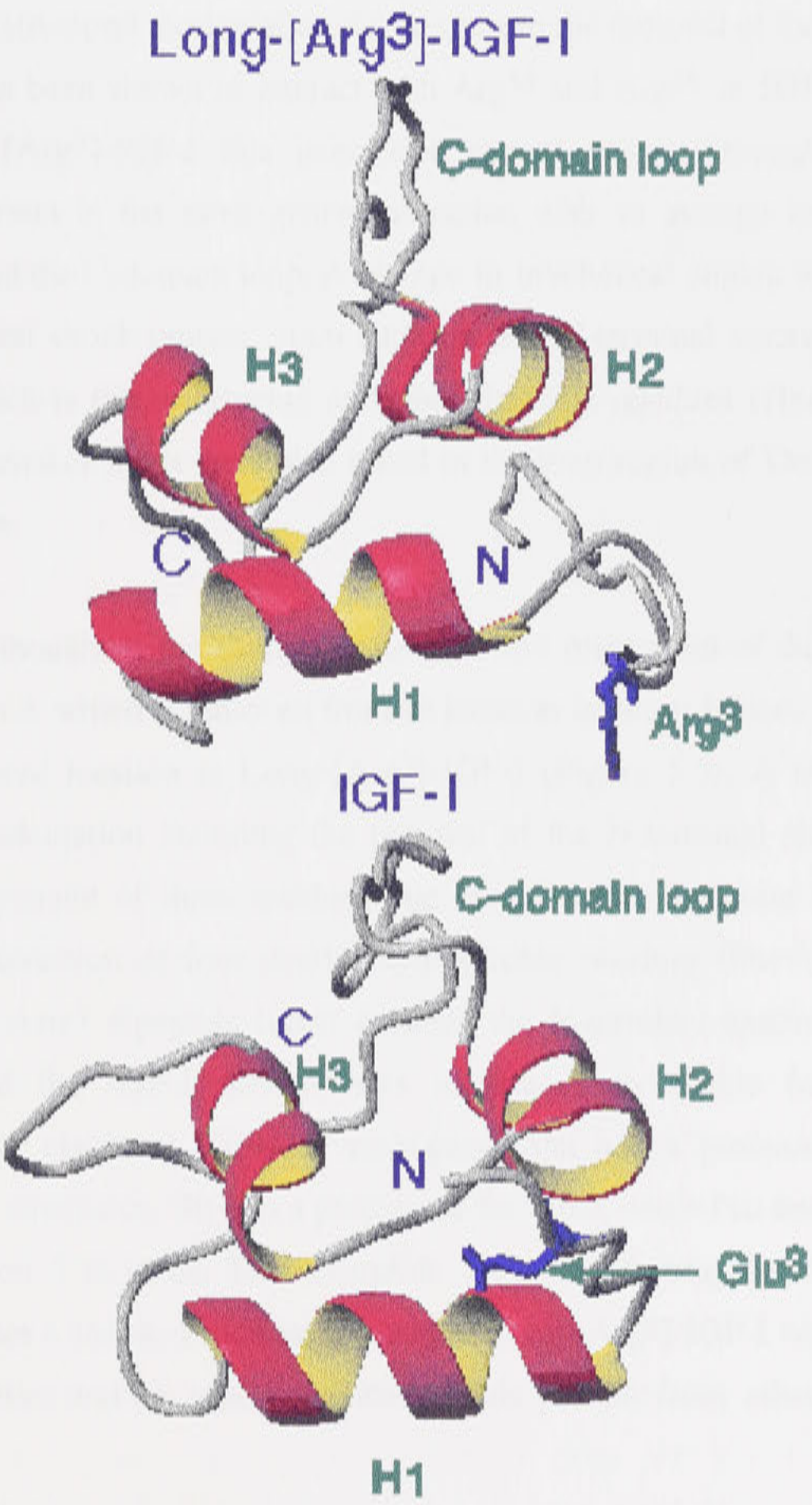

Figure 5.5: Schematic representation of the average NMR structures of Long-[Arg^3]-IGF-I and IGF-I (Cooke et al., 1991). The locations of the sidechains of Arg^3 in Long-[Arg^3]-IGF-I and Glu^3 in IGF-I are shown in blue to indicate the different orientations.

The orientation of the loops in Long-[Arg^3]-IGF-I is not always the same in the various structures, however, the sidechain of Arg^3 is always pointing outwards.

It is unlikely that the orientational shift between helices 1 and 3 is solely a result of additional distance restraints from the ^{15}N-edited experiments because these experiments provide mostly short-range and medium-range distance restraints. Thus, the change is most likely a real structural reorientation resulting from the removal of the free N-terminus at Gly^1 which has been shown to interact with Arg^{55} and Arg^{56} in IGF-I (Cooke et al., 1991). In Long-[Arg^3]-IGF-I this interaction cannot occur, although the N-terminal extension does orient in the same general direction with an average location just above helix 2, helix 3 and the C-domain loop. A change in interhelical angles was also noted for *E. coli* DnaJ, a heat shock protein, upon addition of a C-terminal extension, a conserved "G/F" region which is rich in glycine and phenylalanine residues (Huang et al., 1999). Local conformational changes were also noted in the loop region of DnaJ connecting the two central helices.

Most significant though, is the change in location and orientation of the sidechain of the residue at position 3, which is removed from its location between helices 1 and 3 in IGF-I to a highly exposed location in Long-[Arg^3]-IGF-I (Figure 5.5). A number of factors influence this reorientation including the removal of the N-terminal charge at Gly^1, the advantageous alignment of three residues that favour a turn structure (Asn^{-1}, Gly^1 and Pro^2) and the interaction of four nearby hydrophobic residues (Phe^{-3}, Val^{-2}, Gly^1 and Leu^5). The Val^{-2}-Asn^{-1} dipeptide linker between the N-terminal extension derived from porcine GH and the IGF-I domain was introduced to enable facile cleavage by hydroxylamine for obtaining IGF-I or analogues. Asn has a preference for assuming position 1 in turn structures, Gly has a preference for turns, while Pro has a preference for occupying position 3 in turns. This tripeptide sequence, favourable for assuming turn structures, provides a unique structural element of Long-[Arg^3]-IGF-I which distinguishes the binding affinities and biological activities of this protein from other members of the IGF family.

5.5.2 Implications of Long-[Arg^3]-IGF-I Structure for Binding to IGF Receptors and IGFBPs

Several studies have shown that Arg^{21}, Phe^{23}, Tyr^{24}, Phe^{25} and Val^{44} are crucial for type

1 IGF receptor binding and Cooke et. al. (1991) established that these residues are located in a cleft on one face of IGF-I separating the A- and B- domains from the C-domain. Figure 5.6 shows the relative locations of the N-terminal extension and this cleft in Long-[Arg^3]-IGF-I. It is evident that the N-terminal extension may limit access to Val^{44} and there does appear to be some variation (compared to IGF-I) in the relative positions of Arg^{21}, Phe^{23} and Tyr^{24}, which are crucial to the binding of the type 1 receptor, between the two models.

The three N-terminal residues of IGF-I, in particular Glu^3, are crucial for effective association of IGF-I and IGF-I analogues with the IGFBPs. Mutation of Glu^3 to Arg^3 substantially reduces the affinity of IGF-I analogues for IGFBPs which is logically attributed to an electrostatic effect. The addition of a 13 amino acid extension to the N-terminus of IGF-I and its analogues, [Gly^3]-IGF-I and [Arg^3]-IGF-I further reduces the affinity for the IGFBPs. The amino acid at position 3 is not masked from the IGFBPs by the N-terminal extension, because the "Long" analogues of IGF-I, i.e. Long-IGF-I, Long-[Gly^3]-IGF-I and Long-[Arg^3]-IGF-I, follow the same trend for the binding affinity for IGFBPs as IGF-I, [Gly^3]-IGF-I and [Arg^3]-IGF-I. The solution structure of Long-[Arg^3]-IGF-I agrees with these binding affinity data and indeed shows that Arg^3 in Long-[Arg^3]-IGF-I is even more exposed than Glu^3 in IGF-I (Figure 5.5).

The solution structure of the IGF binding domain of IGFBP-5 has recently been determined and the association of this domain with IGF-II investigated by NMR techniques (Kalus et al., 1998). The authors propose that Glu^6 and Phe^{48} in IGF-II are crucial for the binding of IGF-II to IGFBP-1. By analogy, one can conclude that the spatial location and proximity of the corresponding residues in IGF-I, Glu^3 and Phe^{49}, are important for the binding of IGF-I to IGFBPs. Figure 5.7 shows space-filling models of the location and proximity of Glu^3 and Phe^{49} in IGF-I and Arg^3 and Phe^{49} in Long-[Arg^3]-IGF-I. It is immediately clear that there is a substantial change in the relative positions of these residues, particularly for residue 3, between the two proteins. Thus, the reduced binding of Long-[Arg^3]-IGF-I to IGFBPs may not simply result from the electrostatic effect of the Glu^3 to Arg^3 mutation but also from a spatial effect caused by conformational relocation of the residue at position 3. This would explain the further reduction in affinity for the IGFBPs for the long analogues that is observed in the binding studies. The specific binding determinants of IGF-I analogues for IGFBP-2 are discussed further in chapter 7.

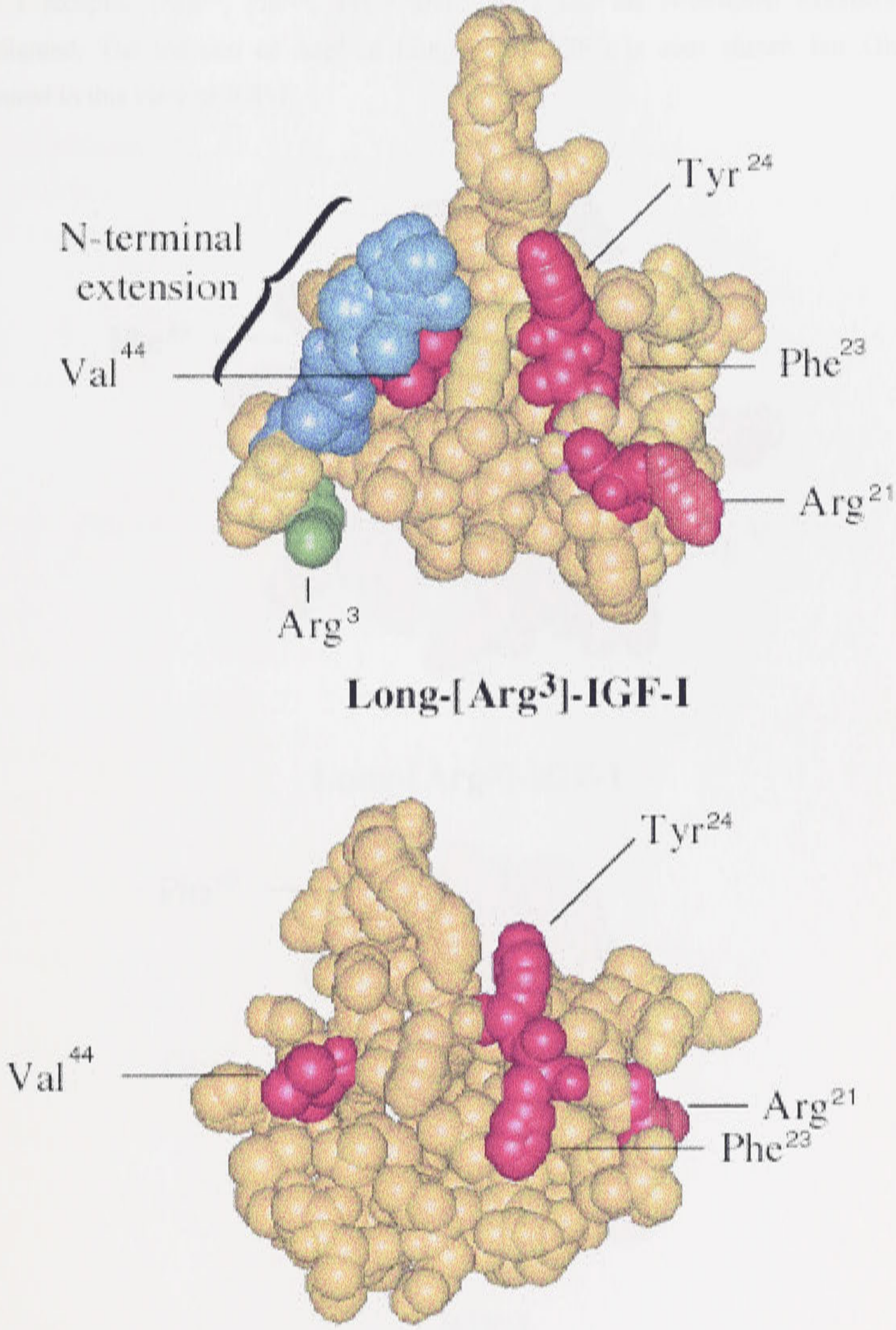

Figure 5.6: A comparison of CPK models of the average NMR structures of Long-[Arg^3]-IGF-I and IGF-I (Cooke et al., 1991). Crucial residues for the interaction with the

type 1 receptor (Arg^{21}, Phe^{23}, Tyr^{24} and Val^{44}) and the N-terminal extension are highlighted. The location of Arg^{3} in Long-[Arg^{3}]-IGF-I is also shown but Glu^{3} is obscured in this view of IGF-I.

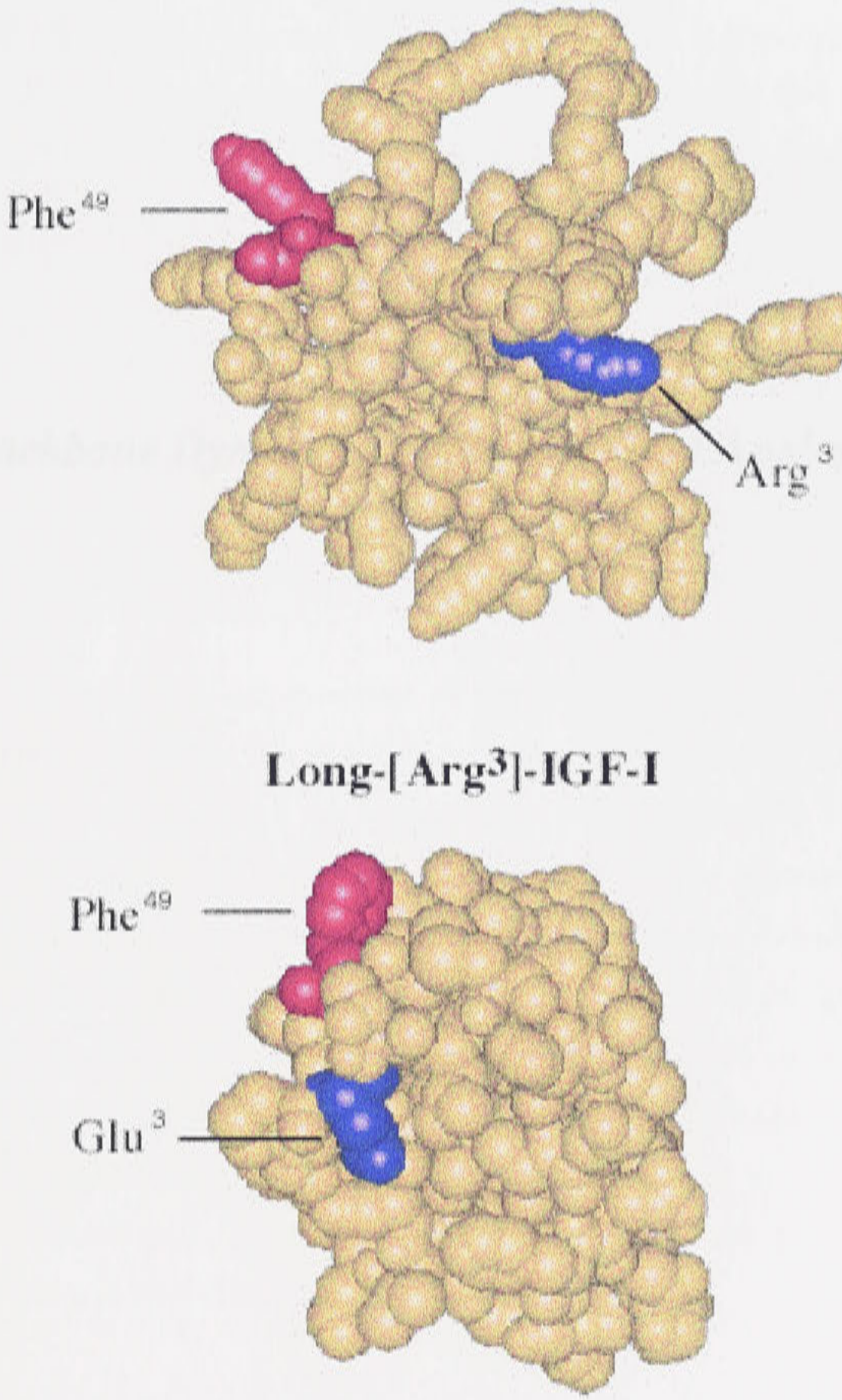

Figure 5.7: A comparison of the CPK models of the average NMR structures of Long-[Arg^{3}]-IGF-I and IGF-I (Cooke et al., 1991) with the residues implicated in the association with IGFBPs (Arg^{3} and Phe^{49} in Long-[Arg^{3}]-IGF-I and Glu^{3} and Phe^{49} in IGF-I) highlighted.

CHAPTER 6

Backbone Dynamics Studies of IGF-I Analogues

6.1 Introduction

Analysis of the backbone dynamics of proteins is a useful adjunct to structural studies because it provides information about the mobility of distinct regions of a protein that cannot be obtained directly from the ^{1}H-^{1}H NOE data. This is particularly important for regions of a protein which have a paucity of ^{1}H-^{1}H NOEs. Both IGF-I analogues, Long-[Arg^{3}]-IGF-I and Long-[Leu^{60}]-IGF-I, have specific regions which have very few ^{1}H-^{1}H NOEs i.e. the N-terminal extension, the C-domain loop and the C-terminus (D-domain) (sections 4.4.1 and 4.4.2). Heteronuclear NMR spin relaxation spectroscopy constitutes a powerful experimental approach for globally characterising conformational dynamics of proteins in solution. The measurement of ^{15}N relaxation rates is particularly useful for obtaining dynamic information since the relaxation of ^{15}N nuclei is governed predominantly by the dipolar interaction with directly bound protons and, to a much smaller extent, by the chemical shift anisotropy mechanism (Allerhand et al., 1971).

Investigations of intramolecular dynamics of proteins by solution NMR spectroscopy involve measurement of the longitudinal relaxation rate constant, R_1, the transverse relaxation rate constant, R_2, and the steady state $\{^{1}H\}^{15}N$ nuclear Overhauser Effect (NOE). These relaxation data are analysed by the model-free approach formulated by Lipari and Szabo (1982a,b). Data can be analysed assuming either isotropic or anisotropic motion. The number of proteins for which backbone and/or sidechain spin relaxation data have been measured in solution is large enough such that comparisons between homologous or related proteins are now possible. Some examples of proteins for which relaxation data have been obtained include interleukin 3 (Feng et al., 1996), the A domain of HMG1 (Broadhurst et al., 1995), hen egg white lysozyme (Buck et al., 1995), and SH2 (Kay et al., 1996) and SH3 (Farrow et al., 1997) domains.

The intramolecular backbone dynamics of both Long-[Arg^{3}]-IGF-I and Long-[Leu^{60}]-IGF-I are presented in this chapter, giving the first example of a member of the IGF family for which this information has been determined.

6.1.1 Theory of Model-free Formalism

The T_1 and T_2 relaxation times and the NOE enhancement of an amide ^{15}N nucleus are dominated by the dipolar interaction of the ^{15}N nucleus with its attached proton and by

chemical shift anisotropy as described by Abragam (1961).

$$1/T_1 = d^2 [J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N) \quad (6.1)$$

$$1/T_2 = (1/2)d^2 [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + (1/6)c^2 [3J(\omega_N) + 4J(0)] \quad (6.2)$$

$$NOE = 1 + (\gamma_N/\gamma_H) d^2 [6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] T_1 \quad (6.3)$$

The constants d^2, and c^2, are defined as

$$d^2 = 0.1 \gamma_H^2 \gamma_N^2 h^2 / (4\pi^2) \langle 1/r^3_{NH} \rangle^2 \quad (6.4)$$

$$c^2 = (2/15) \gamma_N^2 H_o^2 (\sigma_{||} - \sigma_{\perp})^2 \quad (6.5)$$

where γ_H and γ_N are the gyromagnetic ratios of the ^1H and ^{15}N nuclei, respectively, ω_H and ω_N are the ^1H and ^{15}N Larmor frequencies, r_{NH} is the internuclear ^1H-^{15}N distance (1.02 Å), H_o is the magnetic field strength, J is the spectral density function, and the parallel and perpendicular components of the assumed axially symmetrical ^{15}N chemical shift tensor are represented by $\sigma_{||}$ and $\sigma_{\perp}$ respectively. The assumption of an axially symmetric chemical shift tensor has been shown to be valid for peptide bonds with ($\sigma_{||}$ - $\sigma_{\perp}$) = -160 ppm (Hiyama et al., 1988).

The ^{15}N relaxation data are analysed according to the model-free approach developed by Lipari and Szabo (1982a,b). This approach interprets relaxation phenomena in terms of the minimum number of parameters required to describe the overall tumbling motion of a macromolecule and the internal motions of the ^1H-^{15}N bond vector, in the following expression:

$$J(\omega) = S^2\tau_m / (1 + \omega^2\tau_m^2) + (1-S^2)\tau / (1 + \omega^2\tau^2) \quad (6.6)$$

$J(\omega)$ is the spectral density function and ω is the ^{15}N Larmor frequency. The order parameter S^2 describes the degree of spatial restriction of the internal motion of the ^1H-^{15}N bond vector. τ_m is the correlation time of the tumbling motion of the entire molecule.

The effective correlation time of the internal motions is described by τ_e, where $1/\tau = 1/\tau_m + 1/\tau_e$. τ_e is an effective correlation time for the reorientation of the N-H vector due to the internal motion. Isotropic overall tumbling of the molecule is assumed in Equation 6.6.

An extended form of the model-free spectral density function has been developed (Clore et al., 1990 a,b) to describe internal motions that take place on two distinct time scales, differing at least by an order of magnitude. Assuming that the term containing the correlation time describing the faster of the two time scales contributes a negligible amount to the relaxation, the modified spectral density function becomes (Clore et al., 1990b):

$$J(\omega) = S^2\tau_m / (1 + \omega^2\tau_m^2) + (S_f^2 - S^2)\tau / (1 + \omega^2\tau^2) \qquad (6.7)$$

where the order parameter S^2 is expressed as the product of two order parameters characterising the fast and slow internal motions, S_f^2 and S_s^2, respectively. The effective correlation time for the slow internal motions, τ_s, is included using the relationship $1/\tau = 1/\tau_s + 1/\tau_m$.

An additional term, R_{ex}, is required when modelling observed transverse relaxation rates to account for the contributions from processes other than from dipole-dipole and chemical shift anisotropy. In many cases, these contributions are due to conformational exchange averaging (^{15}N exchange broadening during T_2 measurements). The additional term, R_{ex}, may be included to modify the calculated value using the relationship:

$$1/T_2 = 1/T_{2(DD)} + 1/T_{2(CSA)} + R_{ex} \qquad (6.8)$$

in which the DD and CSA subscripts represent contributions from dipole-dipole and chemical shift anisotropy to transverse relaxation (Farrow et al., 1994).

The tumbling of a protein in solution is characterized by a single rotational correlation time only if rotational diffusion is isotropic; an accurate description of anisotropic motion may require up to five correlation times (Woessner, 1962). Preliminary calculations for Long-[Arg^3]-IGF-I revealed a longer than expected global correlation time and suggested some degree of anisotropy. The model-free spectral density function affords a natural separation between isotropic or axially symmetric overall rotational diffusion and the internal motions that are particularly useful for globular proteins. Rotational diffusion anisotropy has a

profound effect on the physics of spin relaxation and on the interpretation of experimental studies of intramolecular dynamics because the spectral density power function depends on the relative orientations of the principal axis systems of the contributing mechanisms and the diffusion tensor (Woessner, 1962). Experimental knowledge of the rotational diffusion tensor is essential for a detailed analysis of intramolecular motions in non-spherical proteins. There are two methods of determining the diffusion tensor: direct fitting of the R_2/R_1 ratios for a set of nuclear spins (Tjandra et al., 1995; Zheng et al., 1995) and analysis of local diffusion coefficients that are derived from relaxation rate constants (Brüschweiler et al., 1995).

For a symmetric top with an axially symmetric diffusion tensor, $D_\perp = D_{xx} = D_{yy}$ and $D_\parallel = D_{zz}$ are the two unique diffusion coefficients, $D_{iso} = (D_\parallel + 2D_\perp)/3$ and;

$$J_i(\omega) = \sum A_{ji} [\tau_j / (1 + \omega^2\tau_j^2)] \qquad (6.9)$$

where $1/\tau_1 = 6D_\perp$, $1/\tau_2 = 5D_\perp + D_\parallel$, $1/\tau_3 = 2D_\perp + 4D_\parallel$, $A_{1i} = (3\cos^2\theta_i - 1)$, $A_{2i} = 3\sin^2\theta_i\cos^2\theta_i$, $A_{3i} = (3/4)\sin^4\theta_i$, and θ_i is the angle between the X-H bond of the ith spin and the unique axis of the principal frame of the diffusion tensor. For the axially symmetric model, the angles θ and ϕ define the orientation of the unique axis of the diffusion tensor frame relative to the arbitrary molecular frame (Lee et al., 1997).

6.2 Materials

6.2.1 Sample

^{15}N-labeled Long-[Arg^3]-IGF-I and ^{15}N-labeled Long-[Leu^{60}]-IGF-I were 1 mM and pH 3.0 in 90% H_2O/10% D_2O. The preparation of samples was described in Section 2.3.4.

6.2.2 Software

Varian VNMR 6.0 (Varian Associates, Palo Alto, CA) and XEASY software were used to measure crosspeak intensities in the T_1, T_2 and NOE experiments. Modelfree 3.1 and Modelfree 4.0 software packages (A.G.Palmer III, Columbia University, Palmer et. al., 1991) were used for the analysis of ^{15}N relaxation data employing the model-free approach of Lipari and Szabo (1982a,b).

6.3 Methods

6.3.1 Heteronuclear NMR Experiments

$R_1(1/T_1)$ and $R_2(1/T_2)$ relaxation rates and $\{^1H\}^{15}N$ steady state NOEs were determined from a series of 2D ^1H-^{15}N correlation spectra at 14.1 Tesla as described by Farrow et al. (1994). The experiments are variations of a sequence originally proposed for the recording of ^1H-^{15}N correlation spectra (Bodenhausen & Ruben, 1980). In order to measure T_1 or T_2 relaxation rates of the heterospin, net magnetisation transfer between the protons and the coupled heteroatom must occur, requiring slightly more complicated pulse sequences than necessary for a normal single quantum heteronuclear correlation experiment. Two refocussed INEPT-type sequences (Morris & Freeman, 1979; Burum & Ernst, 1980) were used to transfer magnetisation from the directly bound protons to the low γ heteronucleus and back to the protons for detection. For the measurement of ^1H-^{15}N NOEs, correlation spectra were recorded separately in the presence and absence of ^1H saturation. Since the longitudinal magnetisation of the heteroatom is to be measured for calculating the heteronuclear NOE, magnetisation must originate on the ^{15}N spin. Magnetisation is subsequently transferred via a refocussed INEPT sequence to the directly coupled H^N proton for observation. The pulse sequences implemented are shown in Appendix 1. ^{15}N T_1 values were measured from spectra recorded with increasing relaxation delays of 5.5, 66, 143, 247.5, 363, 528, 759 and 1144 ms. ^{15}N T_2 values were determined from spectra using a ^{15}N Carr-Purcell-Meiboom-Gill (CPMG) spin-echo sequence with delays of 6, 32, 47, 63, 79, 94, 110 and 142 ms applied during the transverse relaxation period, T, of the T_2 experiment (Carr & Purcell, 1954; Meiboom & Gill, 1958). $\{^1H\}^{15}N$ steady state NOE values were obtained by recording spectra with and without ^1H presaturation of duration 3 s.

$R_1(1/T_1)$ and $R_2(1/T_2)$ relaxation rate constants were determined by non-linear least squares fits of the resonance intensities assuming a single exponential decay. $^1H\{^{15}N\}$ NOE values were obtained from the ratio of intensities of cross peaks in the reference and NOE spectra following the relation $NOE=I_{sat}/I_{nosat}$ where I_{sat} is the intensity of the cross peak when presaturation is applied and I_{nosat} is the intensity of the cross peak in the spectrum with no presaturation.

6.3.2 Initial Estimate of Overall Correlation Time

The overall isotropic correlation time of the protein (τ_m) was estimated from the average T_1/T_2 ratio for residues located in secondary structural elements. An initial estimate of τ_m is usually obtained from the average T_1/T_2 ratio of nuclear spins that do not exhibit slow internal motion and with T_2 values not significantly shortened by chemical or conformational exchange processes (Kay et al., 1989; Clore et al., 1990a). Internal motion results in a reduction in both the T_1/T_2 ratio and the NOE, whereas a significant contribution from exchange processes leads to an increase in the T_1/T_2 ratio. Alternative methods have been used to identify nuclear spins for the estimation of τ_m values. Nuclear spins with slow internal motion can be identified by their small NOE values. Rapid internal motion results in negative NOEs (e.g. in the N- and C-termini of Long-[Arg^3]-IGF-I. A lower limit in the range of 0.6-0.7 for steady state NOE values is frequently adopted for exclusion of nuclear spins in the estimation of τ_m. Nuclear spins having a significantly shortened T_2 value due to chemical or conformational exchange are identified by a T_1/T_2 ratio greater than the average (over the nuclear spins with NOE larger than the preselected value) plus one standard deviation (Kay et al., 1989). Alternatively, nuclear spins exhibiting either slow internal motion or chemical or conformational exchange can be identified exclusively from the T_1/T_2 ratio. For example, only nuclear spins with a T_1/T_2 ratio within one standard deviation of the averaged value are used for the τ_m estimation (Clore et al., 1990a,b), or, a trimmed weighted average value of the T_1/T_2 ratio can be used (Mandel et al., 1995). τ_m is also estimated by making use of structural information, for example, the T_1/T_2 ratios of nuclear spins located in regions that are well defined in the solution structure (Cheng et al., 1993). All of the above methods were used in obtaining an estimate of τ_m for both Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I. For both proteins, the latter method of using the T_1/T_2 ratio of nuclear spins in structured regions was found to result in the least number of nuclear spins requiring optimisation of τ_e and R_{ex}.

6.3.3 Model-free Analysis

The ^{15}N relaxation data were initially analysed using the model-free approach of Lipari and Szabo (1982a,b) with the "Modelfree 3.1" software package (A.G.Palmer III, Columbia University, Palmer et al., 1991; Mandel et al., 1995). Selection of the spectral density functions were based on NOE, T_1 and T_2 values using the protocol of Stone et al. (1992). A preliminary calculation was performed in which τ_m was fixed at the estimated

value, R_{ex} was held at 0, and S^2 and τ_e were optimised for each resonance. In cases where the optimized value of τ_e was non-zero (within 95% confidence limits), τ_e was optimised in the final calculation. Otherwise, τ_e was assumed to be zero in the final calculation. A second preliminary calculation was performed in which τ_m was held at the estimated value and τ_e was held at zero, while S^2 and R_{ex} were optimised for each resonance. In cases where the optimised value of R_{ex} was non-zero, R_{ex} was optimised in the final calculation. Otherwise, R_{ex} was assumed to be zero. A final calculation was performed by optimising τ_m for the whole molecule, S^2 for each resonance, and τ_e and/or R_{ex} for the resonances selected in the two preliminary calculations outlined above.

Subsequently, the relaxation data for Long-[Arg^3]-IGF-I were analysed using Modelfree 4.0 (A.G.Palmer III, Columbia University, Palmer et al., 1998) which optionally utilises an axially symmetric rotational diffusion tensor in the calculations. An initial estimate of this tensor was obtained from the T_1/T_2 ratio and the three-dimensional model of the protein (Chapter 5) using the program, R1R2_diffusion (A.G.Palmer III, Columbia University). The coordinate origin of the three-dimensional structure was translated to the centre of mass using the program, pdbinertia (A.G.Palmer III, Columbia University). Both programs are available at:

http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software/modelfree.html

Long-[Leu^{60}]-IGF-I could not be analysed employing an axially symmetric rotational diffusion tensor model since the three-dimensional structure was not determined for this analogue. Selection of the spectral density functions were based on NOE, T_1 and T_2 values using the protocol of Stone et al. (1993). Briefly, preliminary calculations were performed in which first S^2 and τ_e and then S^2 and R_{ex} were optimised for each amide nitrogen. In the final calculation, τ_m was optimised for the whole molecule, S^2 was optimised for each amide nitrogen and τ_e and R_{ex} were optimised only for those amide nitrogens for which non-zero values for these parameters were found in the preliminary calculations (i.e. outside 95% of set bounds).

6.4 Results

6.4.1 Long-[Arg^3]-IGF-I

A total of 200 relaxation parameters were used as input for the analysis. The T_1, T_2 and T_1/T_2 values were obtained for 74 of the 76 protonated amide nitrogens and are displayed in Figure 6.1. No data were obtained for the nuclear spins of either Cys^{47} or Lys^{68} due to the low intensity of these cross peaks. The mean ^{15}N T_1, T_2 and T_1/T_2 values for residues in helical conformation are 0.74 ± 0.15 s, 0.092 ± 0.046 s and 8.7 ± 2.7 respectively, while the mean T_1, T_2 and T_1/T_2 values for the whole protein are 0.72 ± 0.17 s, 0.13 ± 0.09 s and 6.6 ± 3.2 respectively.

The T_1/T_2 ratio is particularly informative because it can be used to distinguish regions undergoing slow conformational exchange. The mean T_1/T_2 ratio for residues in regions of secondary structure can be used to derive an initial estimate for the overall correlation time for the protein, while deviations from this average identify residues which move in a way that is not entirely correlated with the rest of the molecule. An above average T_1/T_2 ratio indicates T_2 line broadening from conformational exchange. A below average T_1/T_2 ratio suggests that the amide group undergoes motion on a timescale faster than the global correlation time. The value of τ_m estimated in this way of 8.90 ± 0.02 ns, is longer by a factor of two than expected for a globular 9 kDa protein using the Stokes-Einstein relationship. Wagner (1997) provides a plot of τ_m versus molecular weight for a set of 39 proteins or protein complexes. The largest values of τ_m recorded for a protein or protein complex of approximately 10 kDa are 8-9 ns.

The possibility of dimerization of Long-[Arg^3]-IGF-I at the concentration used for the NMR experiments (~1 mM) was investigated and discussed in Chapter 5. At this concentration, the presence of dimer was found to be 20% at most. Thus, the long τ_m value does not arise from large-scale aggregation of Long-[Arg^3]-IGF-I. Anomalously high values of τ_m, estimated from T_1/T_2 ratios may be indicative of anisotropic motion and/or a slow conformational exchange contribution to T_2. The refined model of Long-[Arg^3]-IGF-I indicates that the protein approximates a cylinder with dimensions in the ratio 1.5:1.1:1.0. There is also evidence of conformational exchange in the secondary structural elements that were used to estimate the average T_1/T_2 ratio since an R_{ex} term was required for many of the residues in the α-helical regions of the protein. This is caused in part by slow relative motion of some α-helical regions. Subsequent analysis of

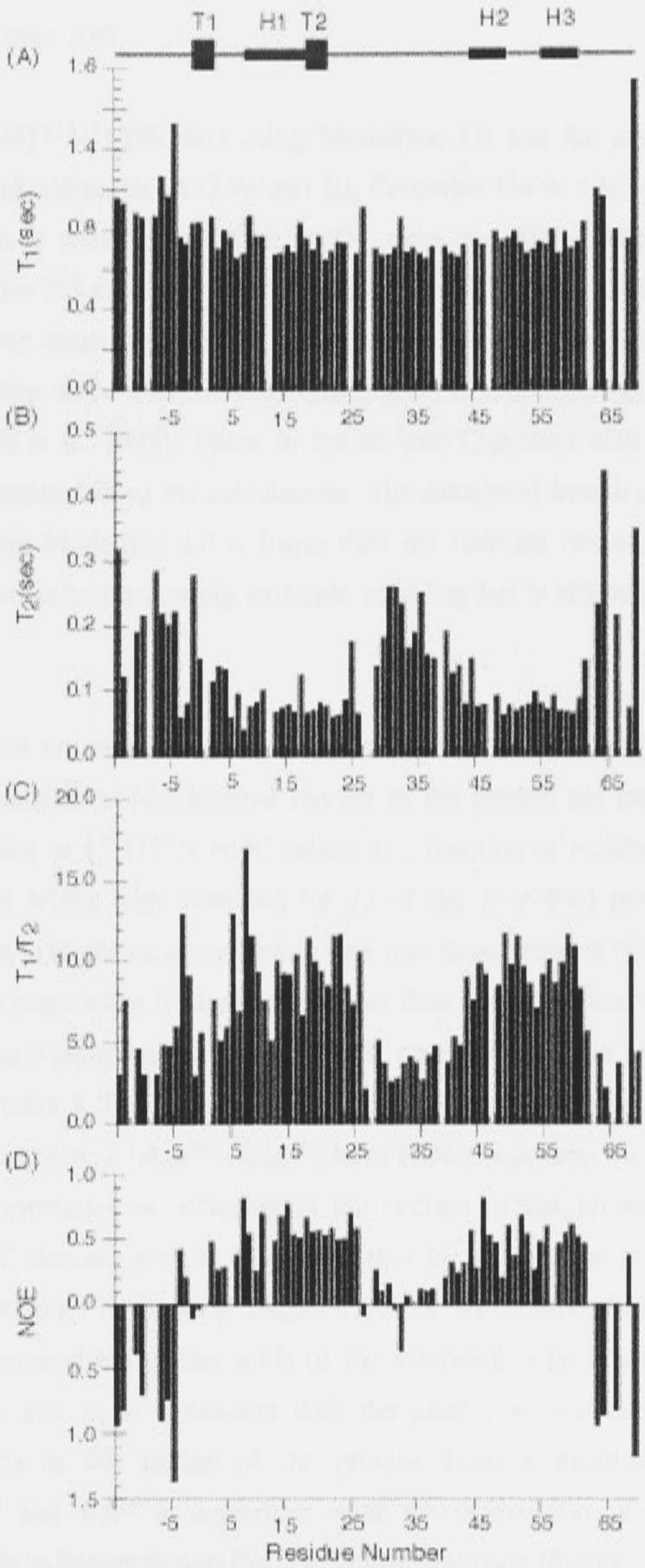

Figure 6.1: Plots of the experimentally determined T_1, T_2, T_1/T_2 and $\{^1H\}^{15}N$ NOE as a function of residue number for Long-[Arg^3]-IGF-I. T_1, T_2, T_1/T_2 and the $\{^1H\}^{15}N$ NOE values are shown in panels (A), (B), (C) and (D) respectively. The error in T_1, T_2 and

NOE is no greater than 10%.

the T_1, T_2 and $\{^1H\}^{15}N$ NOE data using Modelfree 4.0 and the associated programs R1R2_diffusion and pdbinertia (A.G.Palmer III, Columbia University) confirmed that the rotational diffusion is anisotropic with a $D_{||}/D_{\perp}$ ratio of 1.58 ± 0.05 and the axis of diffusion tilted at $\theta = 75^o$ and $\phi = 131^o$ with respect to the magnetic field direction. Spins of residues extending away from the axis of symmetry were excluded when calculating the degree of anisotropy since inclusion of these flexible residues may overestimate the anisotropy (Tjandra et al., 1995). Spins in the N- and C-termini with small or negative NOE values were omitted from the calculations. The calculated overall correlation time of 7.58 ± 0.07 ns from Modelfree 4.0 is lower than the estimate obtained from the T_1/T_2 ratio and the value obtained assuming isotropic tumbling but is still high for a protein of this size.

Heteronuclear NOEs are more sensitive to the effects of internal motion than T_1 or T_2 values, hence the regions of fast internal motion in the protein are more apparent from inspection of the plot of $\{^1H\}^{15}N$ NOE values as a function of residue number (Figure 6.1). Heteronuclear NOEs were obtained for 72 of the 76 (95%) protonated backbone nitrogens. Residues in elements of secondary structure show positive NOEs (average 0.55 ± 0.12) but this average value is significantly less than the theoretical maximum of 0.82 (Kay et al., 1989) indicating that there is substantial motion within the helices, in particular within helix 2 and helix 3. The residues in the poorly defined region of extended structure between helix 1 and helix 2 (Asn^{26} - Gln^{40}) have NOEs near zero indicating substantial conformational reorientation is occurring in this section of the protein. This increased flexibility of this C-domain loop is also supported by an increase in the T_2 relaxation times for several residues in the loop. Negative NOEs are observed for each non-prolyl residue in the N-terminal ten amino acids of the extension, which is indicative of large amplitude motions and is in agreement with the paucity of medium- and long-range homonuclear NOEs in this region of the protein. Positive heteronuclear NOEs are observed at Phe^{-3} and Val^{-2} in agreement with the observation of some long-range homonuclear NOEs in this section of the N-terminal extension (Figure 6.1).

The generalised order parameter, S^2, derived from the relaxation data using the model-free approach (Lipari & Szabo, 1982a,b) describes the degree of spatial restriction of rapid internal motions of the ^1H-^{15}N bond vector of each residue of the protein for which data

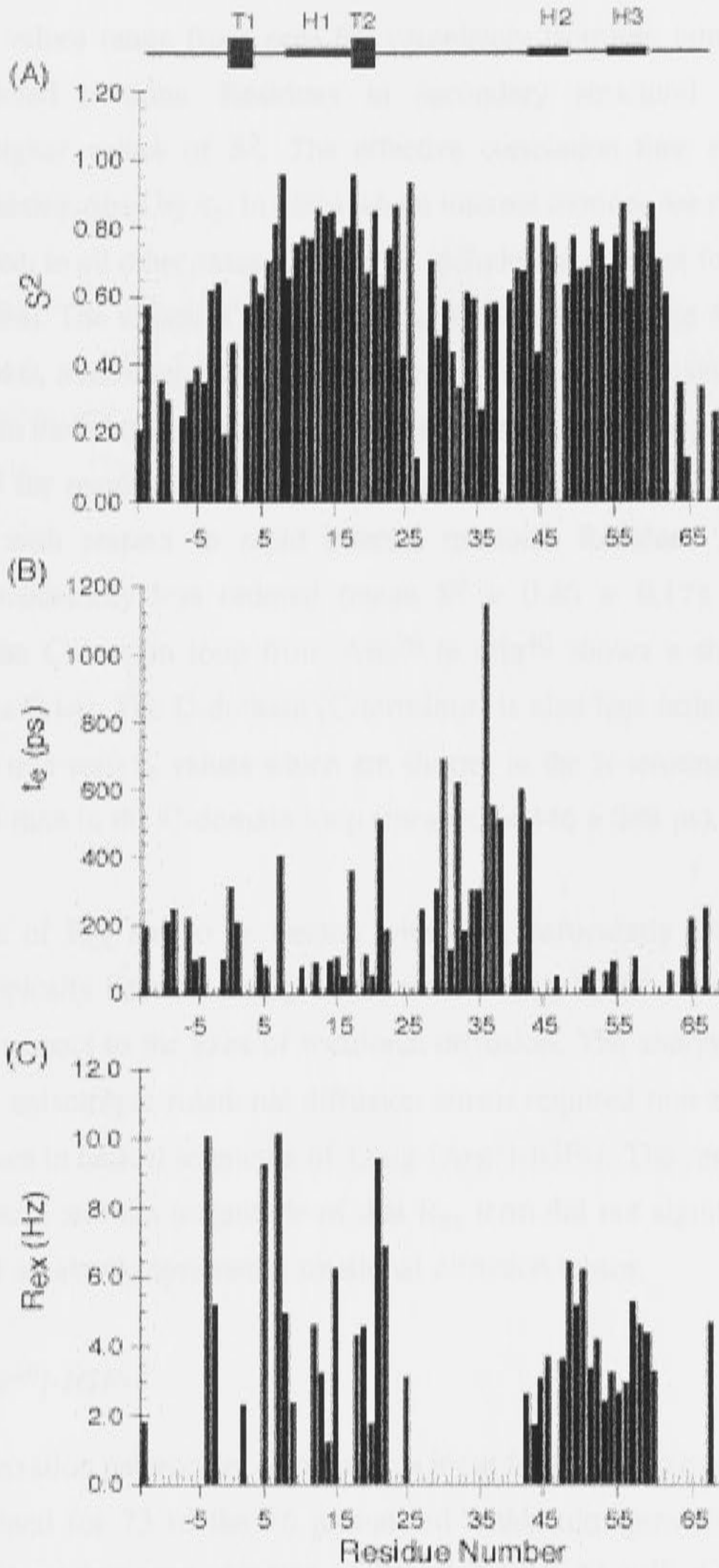

Figure 6.2: Plots of the (A) order parameter (S^2), (B) effective correlation time (τ_e) and (C) exchange contribution (R_{ex}) as a function of residue number for Long-[Arg^3]-IGF-I calculated using Modelfree 4.0 assuming an axially symmetric rotational diffusion model.

are obtained. S^2 values range from zero for completely isotropic tumbling to unity for completely restricted motions. Residues in secondary structural regions therefore generally have higher values of S^2. The effective correlation time resulting from the internal motions is described by τ_e. In cases where internal motions are extremely fast, a τ_e term is not required, in all other cases τ_e must be included to account for internal motions (Farrow et al., 1994). The values of S^2, τ_e and R_{ex} presented in Figure 6.2 were calculated using Modelfree 4.0, assuming an axially symmetric rotational diffusion tensor and are a reasonable guide to the relative flexibility of different sections of Long-[Arg^3]-IGF-I. The mean value of S^2 for residues in helical conformation is 0.81 ± 0.12 indicating a high degree of order with respect to rapid internal motions. Residues in the N-terminal extension are considerably less ordered (mean S^2 = 0.40 ± 0.17). Similarly, the C-terminal part of the C-domain loop from Asn^{26} to Gln^{40} shows a sharp drop in order (mean S^2 = 0.56 ± 0.14). The D-domain (C-terminus) is also less ordered. These flexible regions also have non-zero τ_e values which are shorter in the N-terminal extension (mean τ_e = 186 ± 76 ps) than in the C-domain loop (mean τ_e = 446 ± 288 ps).

The interpretation of R_{ex} has to be treated with care, particularly when the protein is tumbling anisotropically because T_1/T_2 values are sensitive to the orientation of the NH bond vector with respect to the axes of rotational diffusion. The analysis assuming either an isotropic or an anisotropic rotational diffusion tensor required non-zero R_{ex} values for most of the residues in helical segments of Long-[Arg^3]-IGF-I. The residues (nuclei) that required an R_{ex} term and the magnitude of that R_{ex} term did not significantly depend on the assumption of an axially symmetric rotational diffusion tensor.

6.4.2 Long-[Leu^{60}]-IGF-I

A total of 201 relaxation parameters were used as input for model-free analysis. T_1, and T_2 values were obtained for 73 of the 76 protonated amide nitrogens (96%) (Figure 6.3). Residues for which no data were obtained are Met^{-9}, Ser^{-6} and Leu^{10}, due to overlap of the Met^{-9} and Ser^{-6} resonances and the low intensity of the Leu^{10} resonance. The mean ^{15}N T_1, T_2 and T_1/T_2 values for residues in helical conformation are 0.60 ± 0.16 s, 0.09 ± 0.06, and 8.95 ± 4.85 respectively, while the mean T_1, T_2 and T_1/T_2 values for the entire protein are 0.67 ± 0.24, 0.20 ± 0.09 and 6.32 ± 4.51 respectively. The mean T_1/T_2 ratio for residues in regions of secondary structure was used to derive an initial estimate for the overall correlation time of the protein.

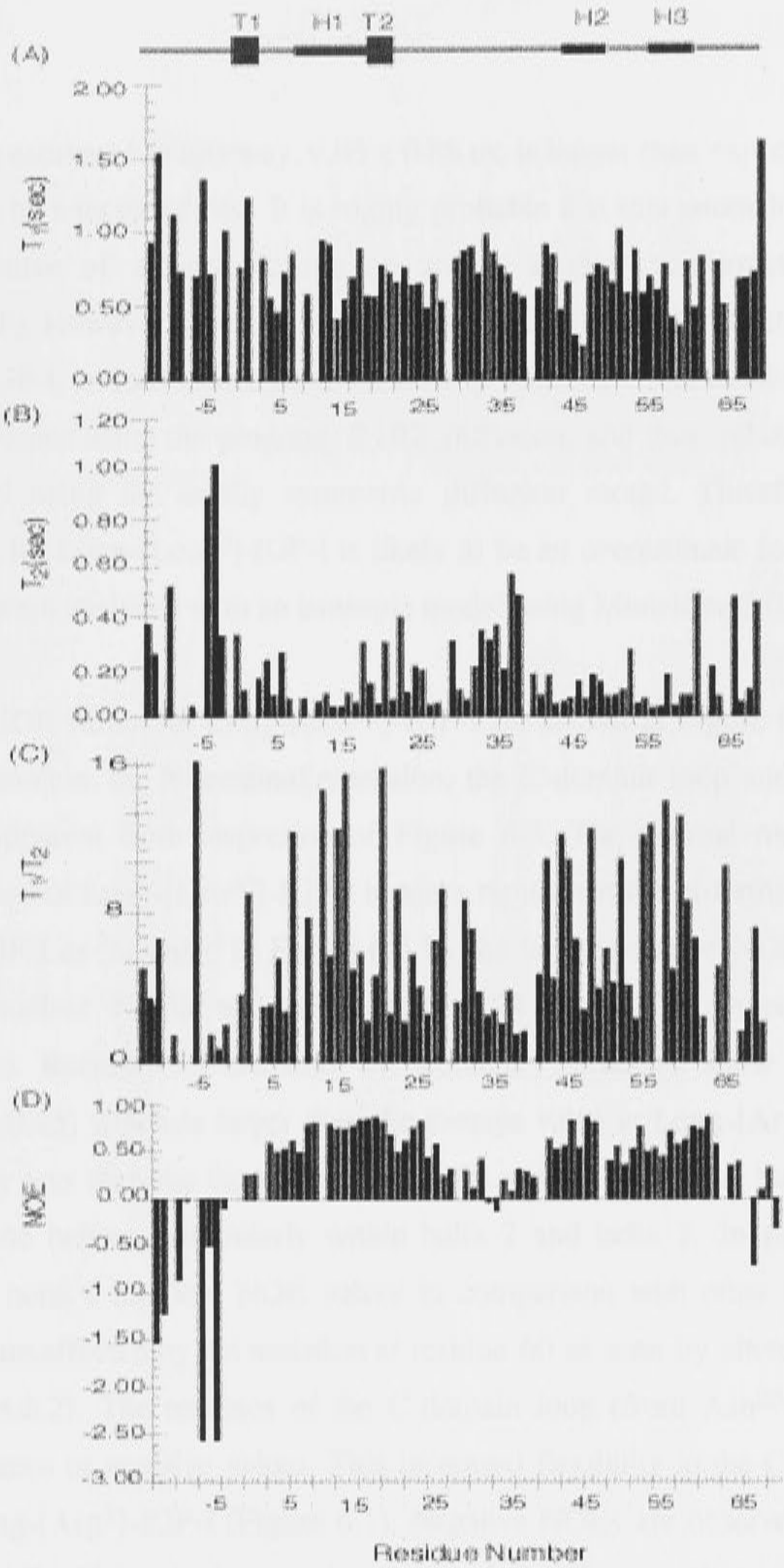

Figure 6.3: Plots of the experimentally determined (A) T_1, (B) T_2, (C) T_1/T_2, and (D) $\{^1H\}$-^{15}N NOE as a function of residue number for Long-[Leu^{60}]-IGF-I. Error in T_1, T_2 and NOE does not exceed 10%.

The value of τ_m estimated in this way, 9.03 ± 0.88 ns, is longer than expected for a 9 kDa globular protein by a factor of two. It is highly probable that this anomalously high value of τ_m is indicative of anisotropic motion and/or a slow conformational exchange contribution to T_2. However, since no three-dimensional structure has been determined for Long-[Leu^{60}]-IGF-I, an initial estimate of the axially symmetric rotational diffusion tensor could not be obtained using the program, R1R2_diffusion, and thus, relaxation data could not be analysed using the axially symmetric diffusion model. Therefore, the overall correlation time for Long-[Leu^{60}]-IGF-I is likely to be an overestimate for this protein as relaxation data were analysed with an isotropic model using Modelfree 3.0.

Heteronuclear NOE values for Long-[Leu^{60}]-IGF-I are shown in Figure 6.3. The regions of fast internal motion, the N-terminal extension, the C-domain loop and the C-terminal extension, are apparent from inspection of Figure 6.3. The internal motion in the N-terminal extension of Long-[Leu^{60}]-IGF-I is more rapid than the N-terminal extension of Long-[Arg^3]-IGF-I as indicated in Figure 6.3 by the larger negative NOE values in this region. Heteronuclear NOEs were obtained for 74 of the 76 protonated backbone nitrogens (97%). Residues in elements of secondary structure show positive NOEs (average 0.69 ± 0.15) which is larger than the average value in Long-[Arg^3]-IGF-I but is still significantly less than the theoretical maximum of 0.82 indicating that there is some motion within the helices, particularly within helix 2 and helix 3. In particular, the C-terminal end of helix 2 has low NOE values in comparison with other helical residues. These residues are affected by the mutation at residue 60 as seen by altered ^{15}N chemical shifts (section 4.2.2). The residues of the C-domain loop (from Asn^{26} to Gln^{40}) have NOEs close to zero or negative values. This increased flexibility in the C-domain loop is also seen in Long-[Arg^3]-IGF-I (Figure 6.1). Negative NOEs are observed for each non-prolyl residue in the N-terminal ten amino acids of the N-terminal extension, Met^{-13} to Leu^{-4}. This is in agreement with the paucity of medium- and long-range homonuclear NOEs in this region of the protein. Phe^{-3} has a heteronuclear NOE of near zero while the segment Val^{-2}-Gly^1 all have small positive NOEs, while in Long-[Arg^3]-IGF-I the NOE observed for Phe^{-3} was positive. This suggests that the mobility of the turn structure involving residues Asn^{-1} to Arg^3 in Long-[Arg^3]-IGF-I is affected by the replacement of Arg^3 by Glu^3 in Long-[Leu^{60}]-IGF-I and this supports the chemical shift data in this region for the two analogues (Chapter 4).

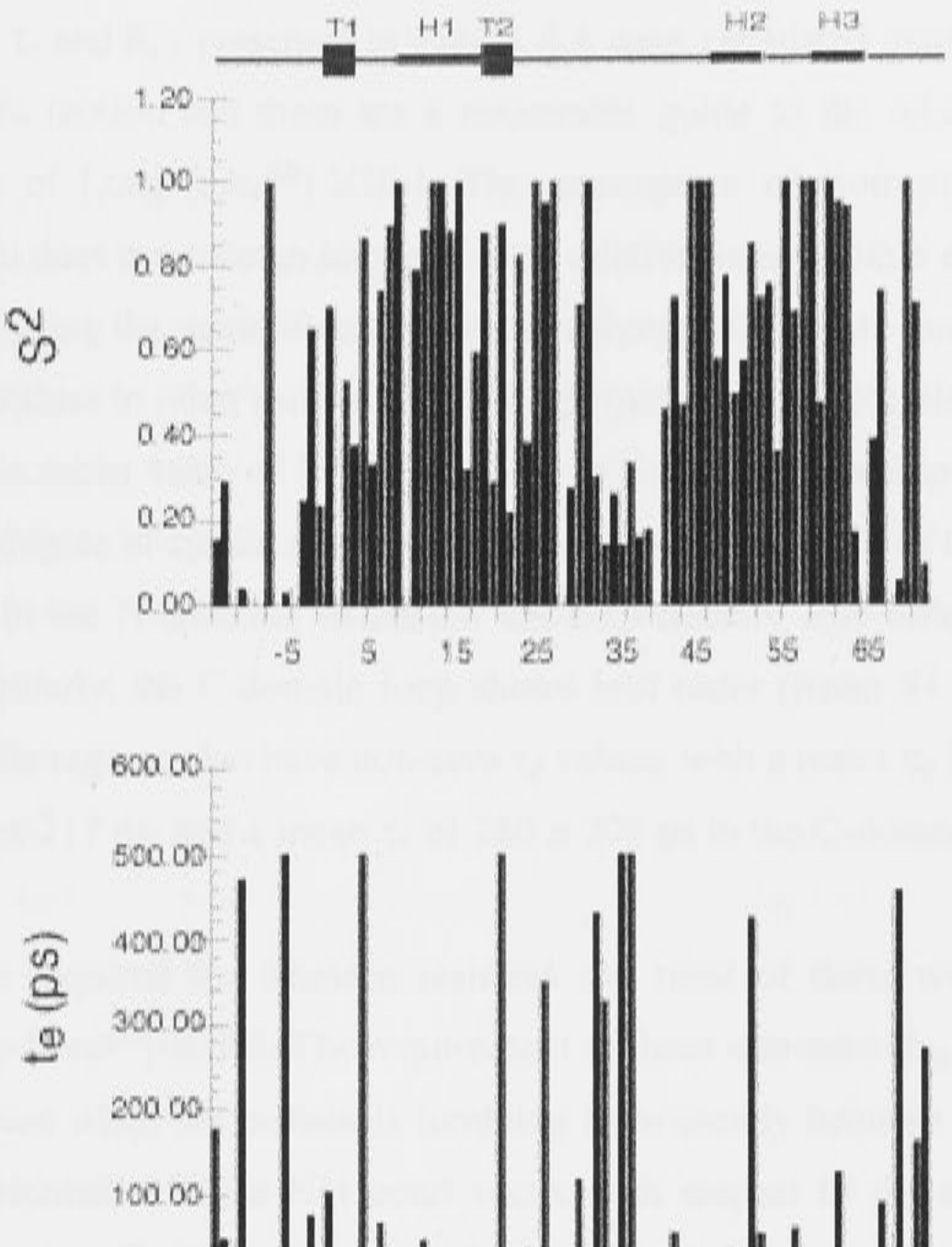

Figure 6.4: Plots of the calculated Modelfree 3.1 parameters, (A) S^2, (B) τ_e and (C) R_{ex} as a function of the residue number for Long-[Leu^{60}]-IGF-I.

The values of S^2, τ_e and R_{ex} presented in Figure 6.4 were calculated using Modelfree 3.1, assuming isotropic motion and these are a reasonable guide to the relative flexibility of different sections of Long-[Leu^{60}]-IGF-I. The assumption of isotropic motion for an anisotropic protein does not substantially affect the relative values within a peptide chain of S^2 and τ_e derived using the model-free formalism, although some care must be taken when comparing these values to other proteins and the optimised overall correlation time will be overestimated. The mean value of S^2 for residues in helical conformation is 0.83 ± 0.21 indicating a high degree of spatial restriction of rapid internal motions of the ^1H-^{15}N bond vector. Residues in the N-terminal extension are considerably less ordered (mean S^2 = 0.29 ± 0.23). Similarly, the C-domain loop shows less order (mean S^2 = 0.34 ± 0.27). These more flexible regions also have non-zero τ_e values, with a mean τ_e in the N-terminal extension of 253 ± 217 ps, and a mean τ_e of 280 ± 209 ps in the C-domain loop.

An R_{ex} term was required for nineteen residues and most of these were in the helical segments of Long-[Leu^{60}]-IGF-I. The requirement of these non-zero R_{ex} values has to be interpreted with care when the protein is tumbling isotropically because T_1/T_2 values are sensitive to the orientation of the NH bond vector with respect to the axes of rotational diffusion. A decrease in T_2 values is observed for these residues, resulting in a T_1/T_2 ratio which is larger than anticipated from the overall correlation time. Table 6.1 presents a summary of spectral density models used to fit T_1, T_2 values and NOE data for both Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I. A large number of τ_e and R_{ex} parameters were necessary to improve the fit for Long-[Leu^{60}]-IGF-I.

Table 6.1. Summary of Spectral Density Models Used to Fit T_1, T_2, and NOE Data for the Two IGF-I Analogues

	Number of Residues	
Parameters Optimised	**Long-[Arg^3]-IGF-I**	**Long-[Leu^{60}]-IGF-I**
S^2	0	30
S^2 and τ_e	46	22
S^2 and R_{ex}	10	13
S^2, τ_e and R_{ex}	18	6
not fitted	9	12

6.5 Discussion

Spectral crowding and low peak intensities were the major cause of difficulties in measuring peak intensities in the T_1, T_2 and NOE spectra. Quantitative analysis of the relaxation data according to the model-free approach is complicated by the presence of 20% Long-[Arg^3]-IGF-I dimer in solution since aggregation systematically increases T_1 and decreases T_2 values. A thorough analysis is only possible when detailed knowledge of the shape and dimensions of the Long-[Arg^3]-IGF-I dimer is known. Qualitative analysis of the relaxation and heteronuclear NOE data in a variety of proteins has provided valuable information about the backbone dynamics. For example, a qualitative analysis of backbone dynamics is adopted for proteins possessing aggregation tendencies, e.g. HIV-1 Nef which exists as approximately 50% monomer and 50% dimer at a concentration of 0.6 mM (Grzesiek et al., 1997). A quantitative analysis using the model-free approach has also been presented in the results employing the most appropriate data-analysis protocol (as discussed below).

Shurr et al. (1994) investigated the effects of anisotropy of rotational diffusion and the extent of dimerisation on the performance of the simple and extended model-free formulas. Fitting to the simple model-free formalism was accomplished by adjusting a single common global correlation time, and also by an alternative protocol, in which the global correlation time was adjusted separately for each nucleus. Anisotropic global rotational diffusion (in the absence of dimerisation) was handled best with the simple model-free formalism adjusting the global correlation time independently for each nucleus, although analysis using the simple model-free formalism adjusting a single overall global correlation time, also gives accurate internal motion parameters. When the degree of dimerisation was between 20% and 90%, none of the model-free approaches tested yields reliable internal motion parameters, although employing the simple model-free formalism with adjustment of a single correlation time, yielded the most accurate internal motion parameters. Thus, in the case of Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I, the relaxation data were analysed using the protocol which yields the most accurate results given that Long-[Arg^3]-IGF-I is subject to both dimerisation and anisotropy of rotational diffusion. Backbone dynamics can be investigated with greater accuracy if ^{13}C-labeled protein is available. The simultaneous analysis of ^{13}C and ^{15}N relaxation data reduces the

bias in the distribution of bond vector orientations used to determine the diffusion tensor (Lee at al., 1997).

The residues of both Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I undergo non-uniform and complicated backbone dynamics. The N-terminal eight residues of the N-terminal extension are not well defined in the solution structure of Long-[Arg^3]-IGF-I (chapter 5) and the order parameters, with an average of less than 0.4, suggest a certain degree of disorder. However, the amide protons of several residues in the N-terminal extension do exhibit slow exchange with the bulk solvent (chapter 4), which suggests that the dynamic disorder is correlated to protect the hydrophobic residues in the N-terminal extension from contact with the aqueous solvent. These eight N-terminal residues in the N-terminal extension also have long T_2 values but the T_1 values are only slightly longer when compared with the rest of the protein, indicating these residues are undergoing low frequency motions. Likewise, the N-terminal eight residues of Long-[Leu^{60}]-IGF-I, with average order parameters of less than 0.3, also display a degree of disorder. The amide protons of several of the same residues of Long-[Leu^{60}]-IGF-I are also slowly exchanging (chapter 4), indicating that the same type of dynamic disorder is occurring in both analogues. In Long-[Arg^3]-IGF-I, the T_2 values shorten markedly at Phe^{-3} and Val^{-2}, which correlates with the observation of several long-range NOEs to these residues. However, in Long-[Leu^{60}]-IGF-I, the T_2 values shorten for the segment Val^{-2} to Gly^1 implying variation in the structure of this region between the two analogues. Most residues in the C-domain loop (Tyr^{24} to Gln^{40}) have S^2 values less than 0.5 and T_2 values greater than 200 ms for both Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I. This degree of disorder for the C-domain loop is consistent with the segment being poorly defined in the solution structure of Long-[Arg^3]-IGF-I. The C-terminal (D-domain) is disordered in both Long-[Leu^{60}]-IGF-I and Long-[Arg^3]-IGF-I. However, a detailed comparison between the two analogues cannot be made due to the paucity of some of the relaxation parameters in this region. Regions of mobility of Long-[Arg^3]-IGF-I are mapped onto the structure of Long-[Arg^3]-IGF-I in Figure 6.5.

The most ordered regions of both Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I are the α-helices. Although the helical regions are highly ordered on the picosecond time scale, the exchange contribution necessary to obtain a good fit to the data suggests that there are slow internal motions that contribute to T_2 (Figures 6.1 and 6.3). This exchange process in all the regions of recognisable secondary structure makes it difficult to estimate τ_m for

both Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I. Although accurate estimates of τ_m are not possible, the trend of S^2 does not vary over a wide range of overall correlation time. Analysis of the data for Long-[Arg^3]-IGF-I assuming either isotropic rotational diffusion or axially symmetric rotational diffusion provides an estimate of τ_m that is longer than expected for a protein of 9 kDa although the anisotropic model estimates a τ_m that is shorter by 20% compared to the isotropic model. Anomalously long values of τ_m, estimated from T_1/T_2 ratios, have been observed in other proteins, for example, in several HMG box proteins (Broadhurst et al., 1995; Jones et al., 1994) and are indicative of anisotropic motion and/or a slow conformational exchange contribution to T_2. Rotational diffusion anisotropies for several proteins have been determined from ^{15}N relaxation measurements (Brüschweiler et al., 1995; Tjandra et al., 1995; Zheng et al., 1995; Mackay et al., 1996). However, due to the constraints imposed by hydrogen bonding in secondary structures, the orientations of the N-H bond vectors are not distributed isotropically in many proteins.

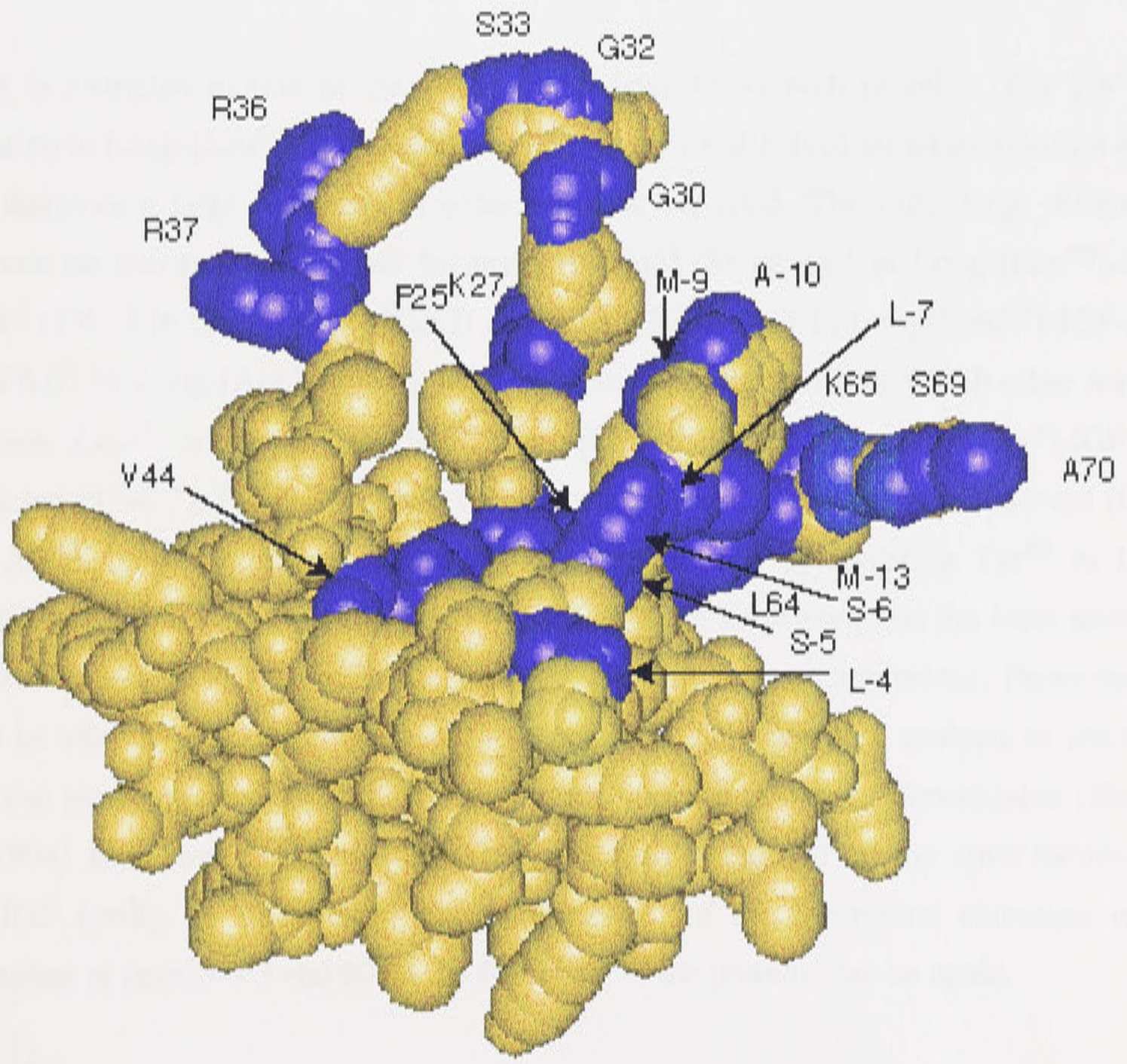

Figure 6.5: Long-[Arg^3]-IGF-I regions with increased motion mapped onto the surface of

Long-[Arg^3]-IGF-I. Residues with low order parameters ($S^2 < 0.6$) are shown in blue.

Correlations between the orientations of the N-H bond vectors compromise the accuracy and precision of the diffusion tensor derived from NMR relaxation measurements (Lee et al., 1997). Lee et al. (1997) demonstrate that simultaneous analysis of $^{13}C^{\alpha}$ and ^{15}N relaxation rate constants improves the reliability of the diffusion tensor determination. Tjandra et al. (1995) observed a higher degree of anisotropy of ubiquitin if the flexibility of the C-terminal residues was not taken into account. The flexibility of the N- and C-terminal residues of Long-[Arg^3]-IGF-I may result in an overestimate of the anisotropy of rotational diffusion. Residues in the N- and C-termini of Long-[Arg^3]-IGF-I with small or negative NOEs were excluded from the calculations when employing the anisotropic rotational diffusion model. Another method of accounting for the flexibility in the N- and C-termini is to calculate the moments of inertia for a number of models of Long-[Arg^3]-IGF-I in which the N-terminal and/or C-terminal region are moved relative to the rest of the model (Hyre & Klevit, 1998).

There is restricted motion in the vicinity of residue 60 in both proteins. The Tyr^{60}Leu mutation in Long-[Leu^{60}]-IGF-I does not disrupt the local helical structure (section 4.4.2) and therefore a large reduction in order was not expected. The only large differences between the two analogues occur for residues Glu^{58} (S^2 of 0.47 in Long-[Leu^{60}]-IGF-I and S^2 of 0.85 in Long-[Arg^3]-IGF-I) and Ala^{62} (S^2 of 0.17 in Long-[Leu^{60}]-IGF-I and S^2 of 0.63 in Long-[Arg^3]-IGF-I). Motion of the N-H bond vector for all other residues between Leu^{57} and Ala^{62} in both Long-[Leu^{60}]-IGF-I and Long-[Arg^3]-IGF-I is restricted (S^2>0.7). The two residues for which a difference in mobility is observed (Glu^{58} and Ala^{62}) both experience a reduction in order induced by mutating Tyr^{60} in Long-[Arg^3]-IGF-I to Leu^{60} in Long-[Leu^{60}]-IGF-I. It is not surprising that the local motion is more restrained in the molecule with the native local amino acid sequence. However, care must be taken in the interpretation of these results since model-free analysis of relaxation data can result in overestimated order parameters in the presence of dimerisation (Shurr et al., 1994). Backbone dynamics studies have not been performed on any other members of the IGF family, so no conclusions on the effect of the N-terminal extension or the mutations at positions 3 and 60 on the flexibility of the proteins can be made.

Both Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I, despite having three well-defined α-helices, are clearly flexible molecules. Long-[Arg^3]-IGF-I retains much of the broad

biological activity of IGF-I except for the reduced affinity for the IGFBPs which also results in enhanced potency in most cultured cell systems. Long-[Leu^{60}]-IGF-I also exhibits reduced binding to IGFBPs due to the 13 amino acid extension and reduced binding to the type 1 IGF receptor, though it displays enhanced potency due to competing endogenous IGF-I from IGFBPs. It has been suggested that a broad specificity of protein interaction with ligands or receptors is correlated with a certain degree of flexibility (Carr et al., 1997). Indeed, it may be necessary for IGF-I to be sufficiently flexible in order to interact with the insulin and the type 1 IGF receptor as well as a family of at least six IGFBPs.

CHAPTER 7

Insulin-Like Growth Factor Binding Protein 2 Interaction with Long-[Leu^{60}]-IGF-I

7.1 Introduction

In vivo, IGFs bind to insulin-like growth factor binding proteins present in extracellular fluids. As a result of this interaction, the half-life of IGFs is prolonged by minimising IGF interaction with receptors. Mutagenesis studies have addressed the localisation of IGFBP binding determinants on the IGFs. The addition of a thirteen amino acid extension at the N-terminus reduces the IGF-I binding to IGFBPs, while residue 3 in IGF-I is critical for high affinity binding of IGF-I to IGFBPs (see section 1.8). In addition, residues Phe^{49}, Arg^{50} and Ser^{51} in IGF-I have been implicated in binding to IGFBPs (Luthi et al., 1992; Bach et al., 1993; Jansson et al., 1998). Dubaquié and Lowman (1999) systematically investigated the binding contribution of each IGF-I amino acid sidechain towards IGFBP-1 and IGFBP-3 by alanine screening mutagenesis. Glu^{3} and Phe^{49} of IGF-I were identified as major specificity determinants for IGFBP-1. No IGF-I sidechain specificity determinant was found for IGFBP-3, instead, the backbone atoms of the N-terminal region of IGF-I were found to be important for binding to IGFBP-3. Clemmons et al. (1992) showed that Glu^{3}, Thr^{4}, and Phe^{49}-Ser^{51} of IGF-I are important for binding of IGF-I to both IGFBP-1 and IGFBP-2. The similar binding pattern of IGF-I for both IGFBP-1 and IGFBP-2 suggests that the structural conformation of IGFBP-2 is similar to IGFBP-1, enabling it to use the same contact points for IGF-I binding.

The NMR chemical shift of a given nucleus is sensitive to its magnetic environment. Perturbations of the electron cloud around a nucleus by an approaching ligand is reflected in a change in the chemical shift. Variations of chemical shifts upon the addition of a ligand are a reliable indication of intermolecular interaction (Omichinski et al., 1993). Jansson et al. (1998) investigated the ^{15}N and ^{13}C shifts of $^{15}N/^{13}C$-labeled IGF-I in the presence and absence of IGFBP-1. Many backbone NMR chemical shifts were found to be affected by IGFBP-1 binding, chiefly in the N-terminal region, C-domain loop and the C-terminal region (D-domain). These include residues Pro^{2}, Glu^{3}, Cys^{6}, Gly^{7}, Pro^{28}, Thr^{29}, Gly^{30}, Gly^{32}, Arg^{36}, Arg^{37}, Gln^{40}, Thr^{41}, Gly^{42}, Ser^{51}, Pro^{63}, Lys^{65}, Pro^{66}, Lys^{68} and Ala^{70}. In addition, three arginine sidechains (Arg^{36}, Arg^{37} and Arg^{50}) were found to participate in IGFBP-1 binding. In the case of Long-[Leu^{60}]-IGF-I, it is of particular interest to determine what effect the N-terminal extension of thirteen amino acids has on the structural binding epitope, since the N-terminal extension decreases the affinity of IGFs for IGFBPs (Francis et al., 1992). The binding affinity of Long-[Leu^{60}]-IGF-I for

IGFBP-2 is investigated at the same pH as for all other NMR work described in this thesis (Chapter 3).

IGFBP-2 binds to Long-[Leu^{60}]-IGF-I and with similar affinity to Long-IGF-I but, with ten times lower affinity than IGF-I itself (Francis et al., 1992; Milner et al., 1999). Thus, the $Tyr^{60}Leu$ mutation does not affect the strength of the binding interaction with IGFBP-2. Since the NMR structural work in this thesis had been performed at pH 3.0, it was necessary to investigate the strength of binding interactions at this pH. The BIAcore™ instrument developed by Pharmacia enables visualisation of macromolecular interactions directly and in "real time" (Fägerstam et al., 1992), and was used to investigate the binding interaction of Long-[Leu^{60}]-IGF-I and IGFBP-2 at pH 3.0. The BIAcore is a biosensor-based instrument that uses surface plasmon resonance as the detection principle. One molecule of the interaction system to be studied is immobilised covalently to a sensor chip, and the other interactant in solution is then passed over the chip. The detection system measures and displays a signal proportional to the mass of protein bound to the surface. In this way, the association phase can be directly visualised as the ligand-containing solution flows over the surface, and the subsequent dissociation is similarly displayed after the flow switches to buffer containing no ligand (Fägerstam et al., 1992).

Chemical shift mapping experiments, that exploit the sensitivity of the NMR chemical shifts to the local environments, have found recent widespread application for defining the binding epitopes of proteins particularly if the protein is labeled with stable isotopes (Grzesiek, et al., 1997; Jansson et al., 1998; Kalus et al., 1998). Recently, Jansson et al. (1998) used the ^{13}C and ^{15}N chemical shift perturbations of twenty residues on one face of IGF-I to define the binding interface with IGFBP-1. Kalus et al. (1998) performed similar experiments but in this case a mini-IGFBP-5 was labeled with ^{15}N and ^{13}C and the complexed ligand was IGF-II.

In order to gain a more detailed understanding of the nature of interactions between the insulin-like growth factors and their binding proteins, the ^{15}N chemical shifts of Long-[Leu^{60}]-IGF-I in the presence and absence of IGFBP-2 have been compared to identify the IGFBP-2 structural binding epitope.

7.2 Materials and Methods

7.2.1 Binding Protein

Insulin-Like Growth Factor Binding Protein-2 (IGFBP-2) (10 mg) was purchased from GroPep Pty. Ltd. (Adelaide, Australia). IGFBP-2 was produced following a procedure adapted from Szabo et al. (1988).

7.2.2 BIAcore Binding Studies

Binding studies were performed on a BIAcore 2000 with BIAevaluation software (version 3.0.2). BIAcore analyses were performed with Long-[Leu^{60}]-IGF-I as the immobilised ligand. All analyses were performed at 25 °C and followed the protocol of Hobba et al. (1998). Covalent attachment of Long-[Leu^{60}]-IGF-I to the CM5 biosensor chip was achieved by the amine coupling method via the ε-NH_3^+ groups of lysine (Löfås and Johnson, 1991). Long-[Leu^{60}]-IGF-I (226 µg/ml in 50 mM sodium acetate, pH 7.4) was injected onto the activated CM5 surface at 5 µl/min with HBS (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, pH 7.4) as the running buffer. Long-[Leu^{60}]-IGF-I was immobilised such that the final resonance value was approximately 400 above the resonance value of the activated but underivatized chip. Residual binding sites were quenched with ethanolamine.

In order to estimate the dissociation constant (K_D) prior to a kinetic study, two or three concentrations of IGFBP-2 were injected for 5 min at a flow rate of 40 µl/min with HBS as the running buffer. For a kinetic study, ten concentrations were chosen based on the trial: 2, 5, 8, 11, 15, 20, 30, 50, 100 and 150 nM. The dissociation phase, initiated by switching from the stream of IGFBP-2 to HBS, was carried out over a period of 10 min. The Long-[Leu^{60}]-IGF-I surfaces were regenerated by a 90 second injection of 0.1 M HCl to dissociate bound IGFBP-2. The binding procedure was repeated at pH 3.0 (10 mM citric acid, 150 mM NaCl and 3 mM EDTA) and pH 4.0 (10 mM formate, 150 mM NaCl and 3 mM EDTA). Trials were performed at both pH 3.0 and pH 4.0 with IGFBP-2 injected for 5 min at a flow rate of 40 µl/min. Concentrations used for the kinetic study were based on trials in which two or three concentrations were used to determine the approximate K_D, and concentrations were chosen around this for the kinetic study. The concentrations chosen for the kinetic study at both pH 3.0 and pH 4.0 were 5, 10, 25, 50, 75, 100, 200 and 400 nM. The apparent association and dissociation rates of IGFBP-2

were derived by fitting the experimental data to a one-site association and dissociation model with the BIAevaluation software (version 3.0.2).

7.2.3 Formation of the Long-[Leu^{60}]-IGF-I/IGFBP-2 Complex for NMR Spectroscopy

1 mg of IGFBP-2 was dissolved in 150 µl 10 mM citrate buffer, pH 3.0. This solution was added to 325 µl of 0.34 mM Long-[Leu^{60}]-IGF-I at pH 3.0, to give a Long-[Leu^{60}]-IGF-I concentration of 0.23 mM in 475 µl and a IGFBP-2 concentration of 0.066 mM. The IGFBP-2 concentration was at a 0.28:1.0 molar ratio relative to Long-[Leu^{60}]-IGF-I. IGFBP-2 was then progressively added to the solution as a dry powder to obtain concentrations of IGFBP-2 relative to Long-[Leu^{60}]-IGF-I of 48%, 83% and 100%. No precipitation of either Long-[Leu^{60}]-IGF-I or IGFBP-2 was observed at pH 3.0.

7.2.4 NMR Experimental

The Long-[Leu^{60}]-IGF-I/IGFBP-2 complex was studied by NMR spectroscopy at pH 3.0. 2D ^{15}N-HSQC spectra (Chapter 3) were successively recorded for mixtures of Long-[Leu^{60}]-IGF-I and IGFBP-2 with IGFBP-2 concentrations of 0:1.0, 0.28:1.0, 0.48:1.0, and 0.83:1.0 molar ratio relative to Long-[Leu^{60}]-IGF-I. The $^{1}H_N$ and ^{15}N chemical shifts of all cross-peaks were measured at each IGFBP-2 concentration.

7.3 Results

7.3.1 BIAcore Analysis

Kinetic analyses of the association and dissociation of IGFBP-2 with immobilised Long-[Leu^{60}]-IGF-I at pH 3.0, pH 4.0 and pH 7.4 were performed in the BIAcore. The purpose of performing these binding studies was to determine if the binding affinity of Long-[Leu^{60}]-IGF-I for IGFBP-2 was sufficiently strong to perform NMR studies of the complex. This was confirmed, with a K_D in the nanomolar range at pH 3.0.

Binding experiments at pH 7.4 were performed to enable comparison with a previously obtained dissociation constant (K_D 10.3 ± 3.0 nM) for IGFBP-2 binding to Long-[Leu^{60}]-IGF-I (Milner et al., submitted for publication) and to confirm that results were reproducible. The dissociation constant obtained for Long-[Leu^{60}]-IGF-I of 14.3 ± 0.1

nM was close to the value obtained by Milner et al. (submitted for publication). Kinetic analyses for Long-[Leu^{60}]-IGF-I binding to IGFBP-2 at pH 7.4 are shown in Figure 7.1. The interaction between IGFBP-2 and Long-[Leu^{60}]-IGF-I at pH 7.4 was difficult to resolve to a single binding site. Similar behaviour was also found by Hobba et al. (1998) when using IGF-I and IGF-II as the immobilised ligands and monitoring their interactions with IGFBP-2 and IGFBP-2 mutants.

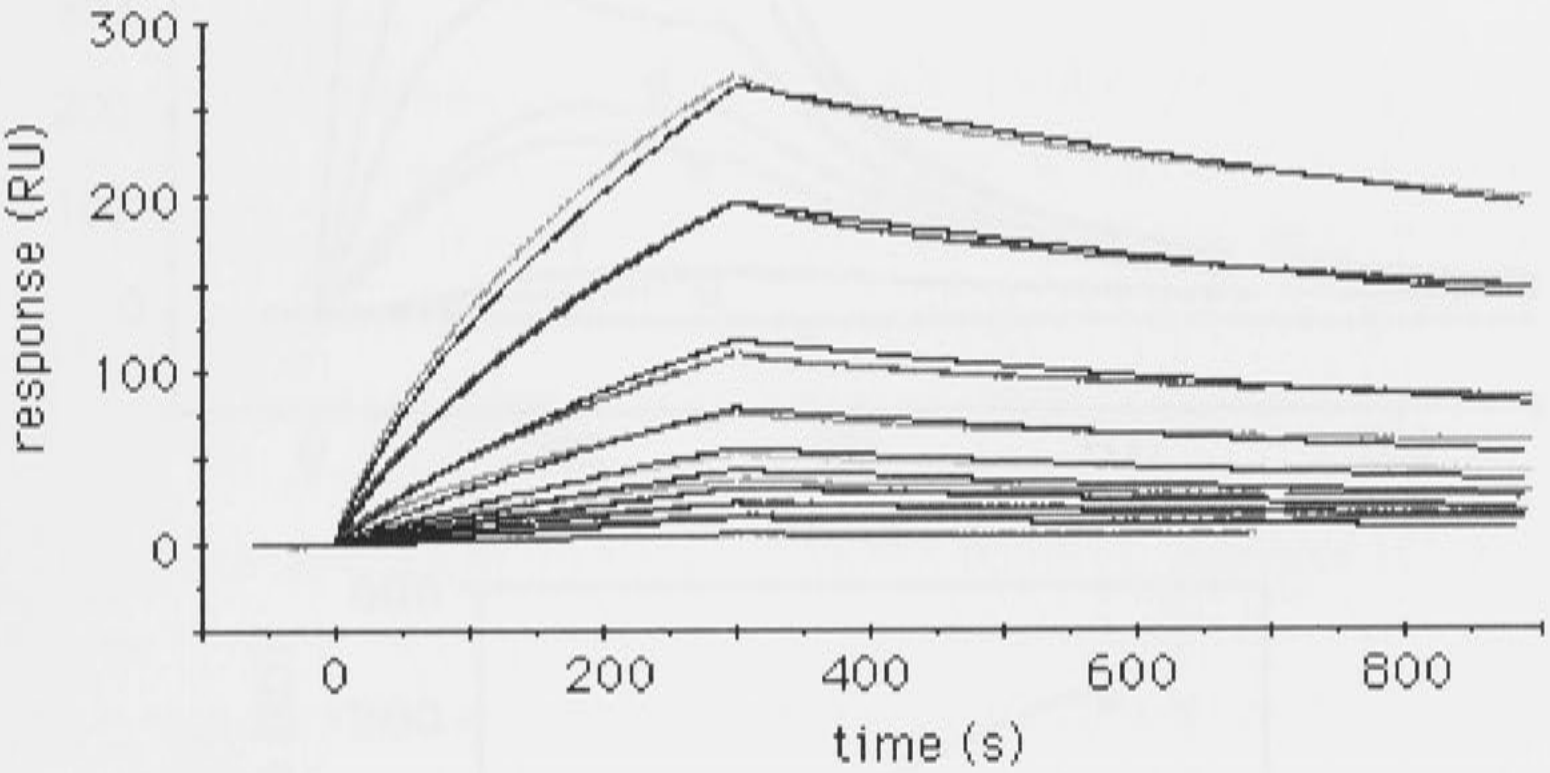

Figure 7.1: Association and dissociation curves for IGFBP-2 binding to immobilized Long-[Leu^{60}]-IGF-I at pH 7.4. The plot shows the relative response in resonance units (RU) plotted against time for different IGFBP-2 concentrations; 2 nM cyan, 5 nM red, 8 nM dark green, 11 nM light blue, 15 nM pink, 20 nM yellow, 30 nM light green, 50 nM dark blue, 100 nM burgundy, 150 nM grey.

The binding curves for IGFBP-2 to Long-[Leu^{60}]-IGF-I at pH 3.0 indicated that saturation of binding sites occurred prior to the end of the association phase. However, for four of the chosen concentrations, the response decreased prior to initiating dissociation by switching from the IGFBP-2 stream to the buffer to initiate dissociation. The analyte dissociation rate was calculated by plotting response at the end of the association phase (approximately 300 s) against concentration for a point on each of the curves during the association phase (Figure 7.2). The K_D measured in this way was 200 ± 100 nM, an approximately ten-fold higher value than at pH 7.4. Binding data for IGFBP-2 to Long-[Leu^{60}]-IGF-I at pH 4.0 was irregular and the data could not be fitted to a model to derive the association and dissociation constants. Jansson et al. (1998) reported that between pH 3.4 and pH 4.8 (the pI of IGFBP-1) IGFBP-1 precipitates, providing an explanation for

this behaviour. However, the pI of IGFBP-2 has been calculated as 7.15 using "Pepstats" on ANGIS (Australian National Genomic Information Service, 1998).

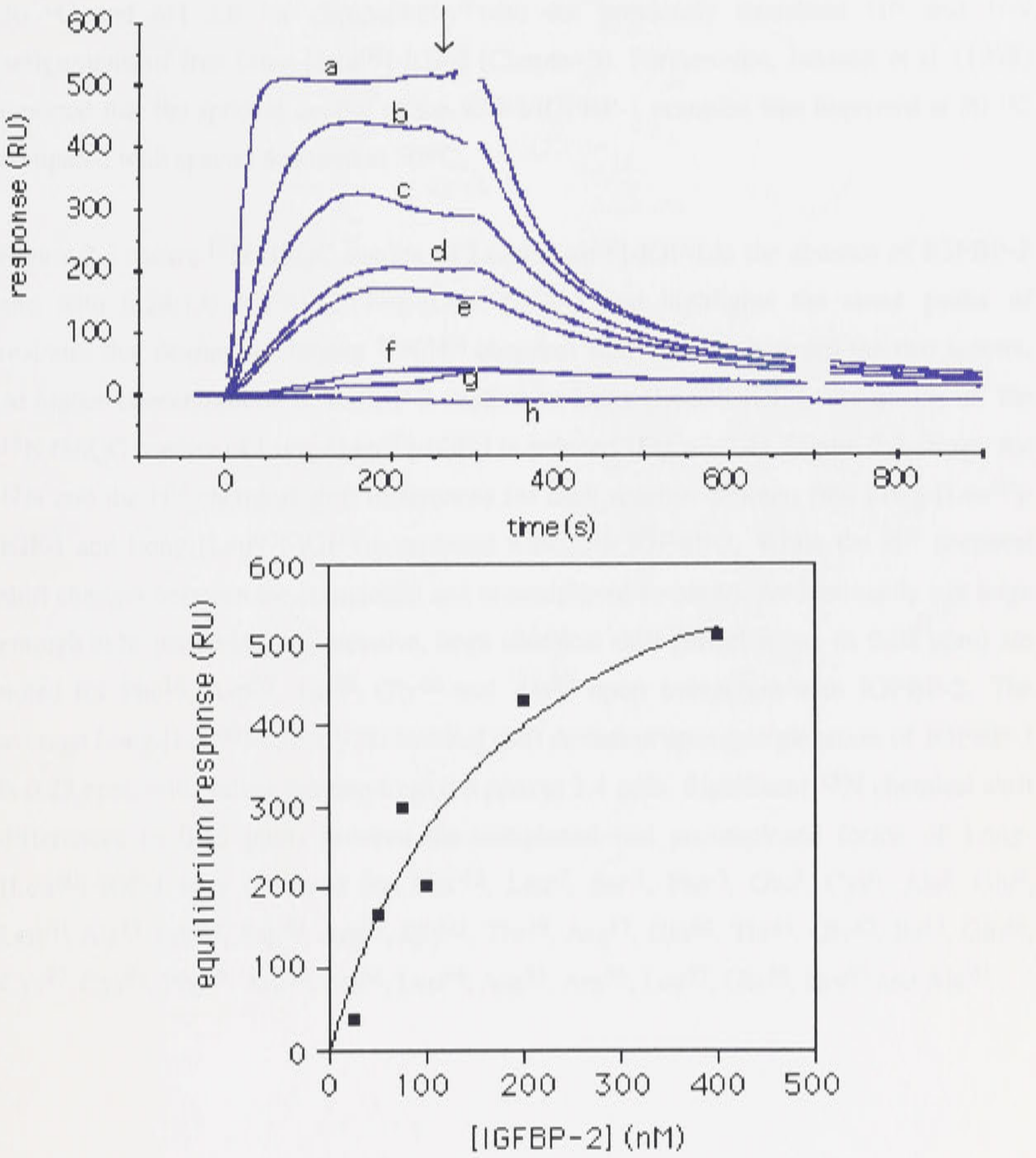

Figure 7.2: BIAcore analysis of IGFBP-2 binding to Long-[Leu^{60}]-IGF-I at pH 3.0. (A) The plot shows the relative response in resonance unit (RU) plotted against time for different concentrations of IGFBP-2; (a) 400 nM, (b) 200 nM, (c) 100 nM, (d) 75 nM, (e) 50 nM, (f) 25 nM, (g) 10 nM, (h) 5 nM, and (B) shows the steady state affinity curve for IGFBP-2 binding to Long-[Leu^{60}]-IGF-I at pH 3.0.

7.3.2 Mapping the Long-[Leu^{60}]-IGF-I Binding Site

2D ^{15}N-HSQC spectra of Long-[Leu^{60}]-IGF-I complexed to IGFBP-2 were acquired at 30 ^{o}C and pH 3.0 for compatibility with the previously described H^N and ^{15}N assignments of free Long-[Leu^{60}]-IGF-I (Chapter 3). Furthermore, Jansson et al. (1998) reported that the spectral quality of the IGF-I/IGFBP-1 complex was improved at 30 ^{o}C compared with spectra acquired at 50 ^{o}C.

Figure 7.3 shows ^{15}N HSQC spectra of Long-[Leu^{60}]-IGF-I in the absence of IGFBP-2 and with 0.28:1.0 IGFBP-2:Long-[Leu^{60}]-IGF-I and highlights the cross peaks of residues that display the largest ^{15}N/H^N chemical shift changes between the two spectra. At higher concentrations of IGFBP-2 relative to Long-[Leu^{60}]-IGF-I, the quality of the ^{15}N HSQC spectra of Long-[Leu^{60}]-IGF-I is reduced (Figure 7.4). Figure 7.5 shows the ^{15}N and the H^N chemical shift differences for each residue between free Long-[Leu^{60}]-IGF-I and Long-[Leu^{60}]-IGF-I complexed with 25% IGFBP-2. While the H^N chemical shift changes between the complexed and uncomplexed forms are predominantly not large enough to be particularly informative, large chemical shift perturbations (≥ 0.08 ppm) are noted for Phe^{16}, Asp^{20}, Thr^{29}, Gly^{42} and Ala^{70} upon interaction with IGFBP-2. The average Long-[Leu^{60}]-IGF-I ^{15}N chemical shift deviation upon complexation of IGFBP-2 is 0.23 ppm, with values ranging from 0.0 ppm to 2.4 ppm. Significant ^{15}N chemical shift differences (> 0.15 ppm) between the complexed and uncomplexed forms of Long-[Leu^{60}]-IGF-I were observed for Met^{-13}, Leu^{-7}, Ser^{-5}, Phe^{-3}, Glu^{3}, Cys^{6}, Ala^{8}, Glu^{9}, Leu^{10}, Ala^{13}, Leu^{14}, Asp^{20}, Arg^{21}, Gly^{22}, Thr^{29}, Arg^{37}, Gln^{40}, Thr^{41}, Gly^{42}, Ile^{43}, Glu^{46}, Cys^{47}, Cys^{48}, Phe^{49}, Arg^{50}, Ser^{51}, Leu^{54}, Arg^{55}, Arg^{56}, Leu^{57}, Glu^{58}, Lys^{65} and Ala^{70}.

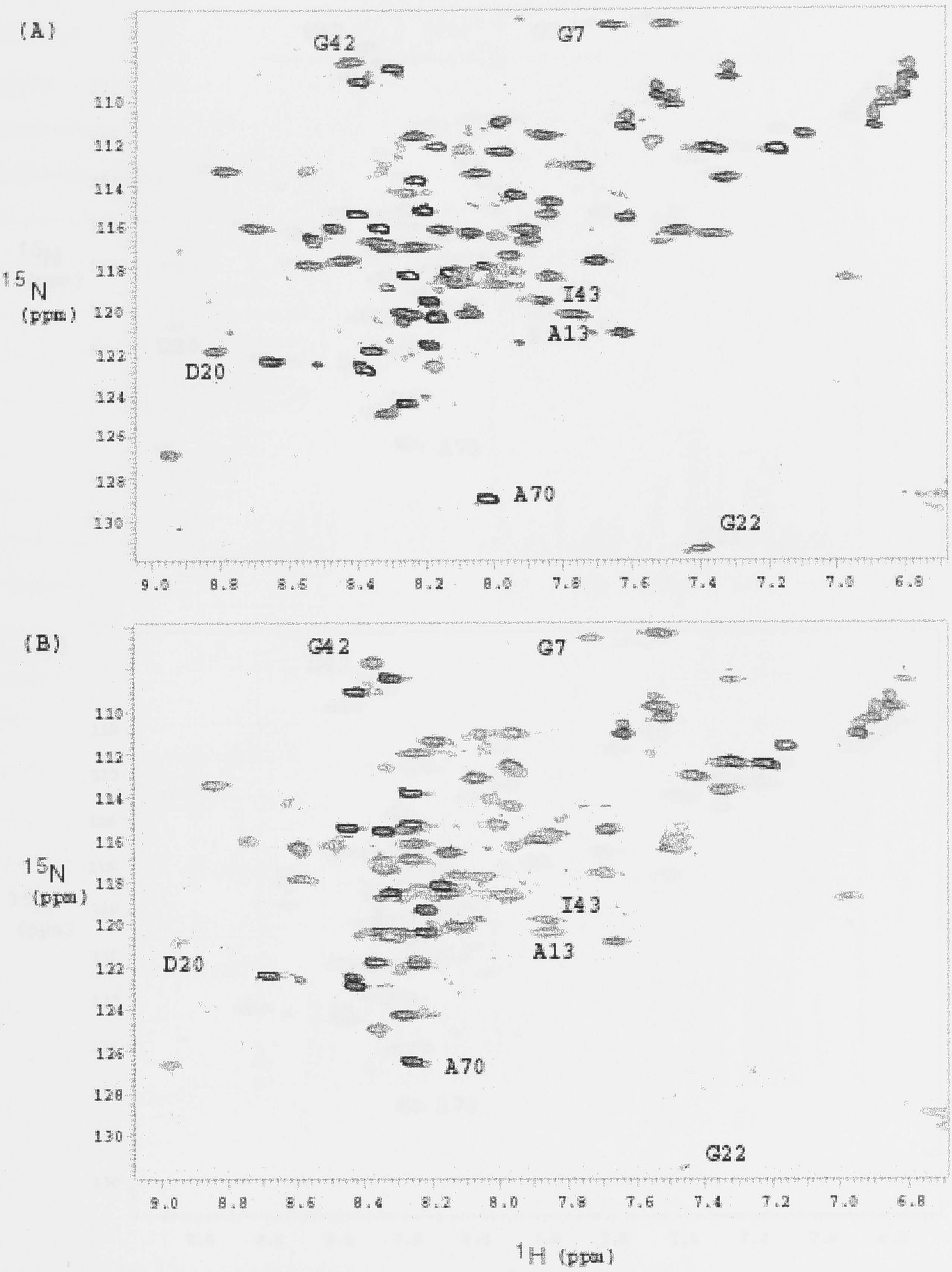

Figure 7.3: 2D ^{15}N-HSQC spectra of Long-[Leu^{60}]-IGF-I; (A) without IGFBP-2, and (B) with 0.25:1.0 molar ratio IGFBP-2:Long-[Leu^{60}]-IGF-I. Some residues which experience significant ^{15}N chemical shift differences upon IGFBP-2 binding are labeled.

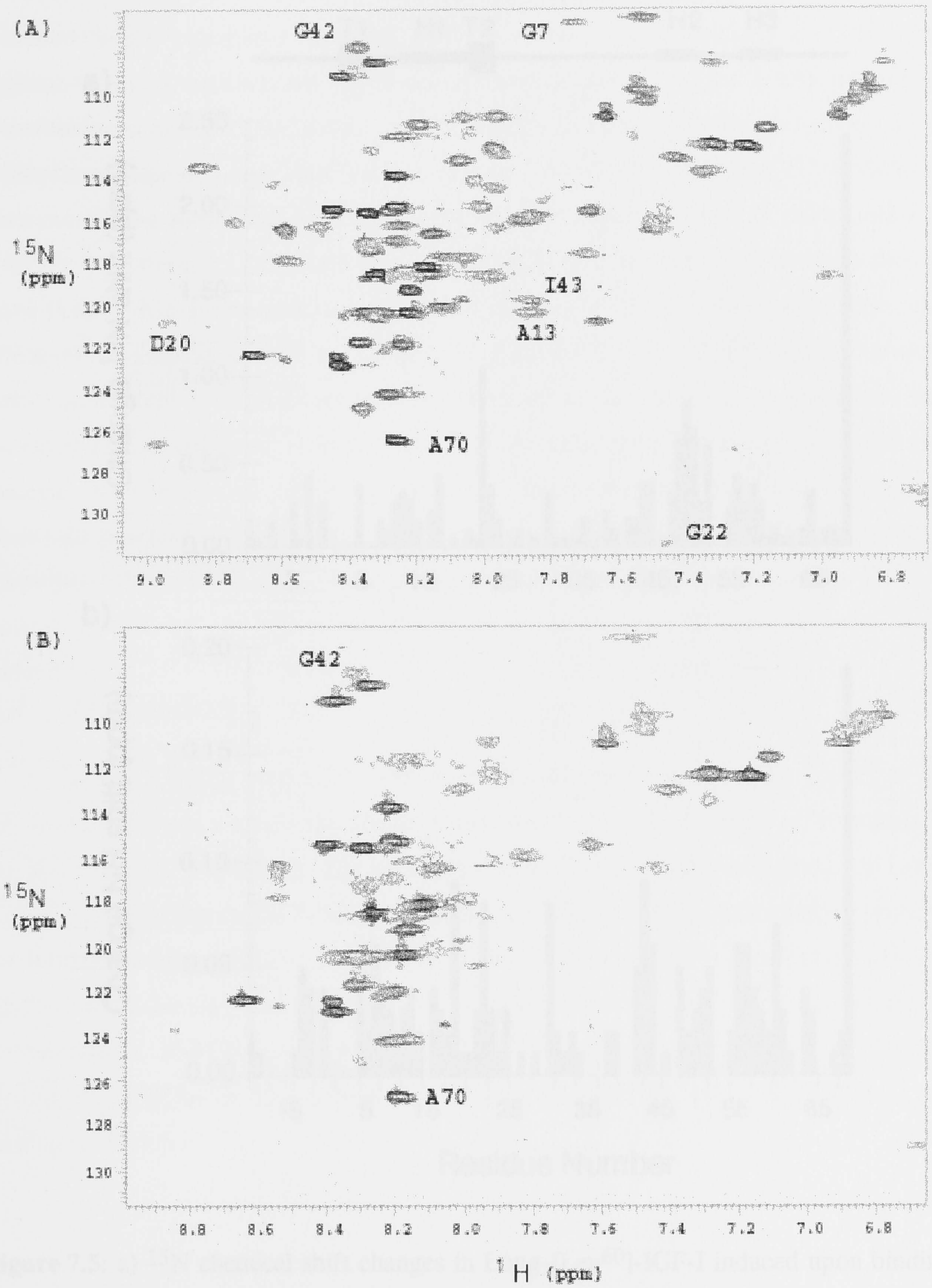

Figure 7.4: 2D ^{15}N-HSQC spectra of Long-[Leu^{60}]-IGF-I with; (A) 0.25:1.0 molar ratio IGFBP-2:Long-[Leu^{60}]-IGF-I, and (B) 0.83:1.0 molar ratio IGFBP-2:Long-[Leu^{60}]-IGF-I. Comparison of the 2D ^{15}N-HSQC spectra of Long-[Leu^{60}]-IGF-I shows increased line broadening and decreased spectral resolution upon addition of IGFBP-2.

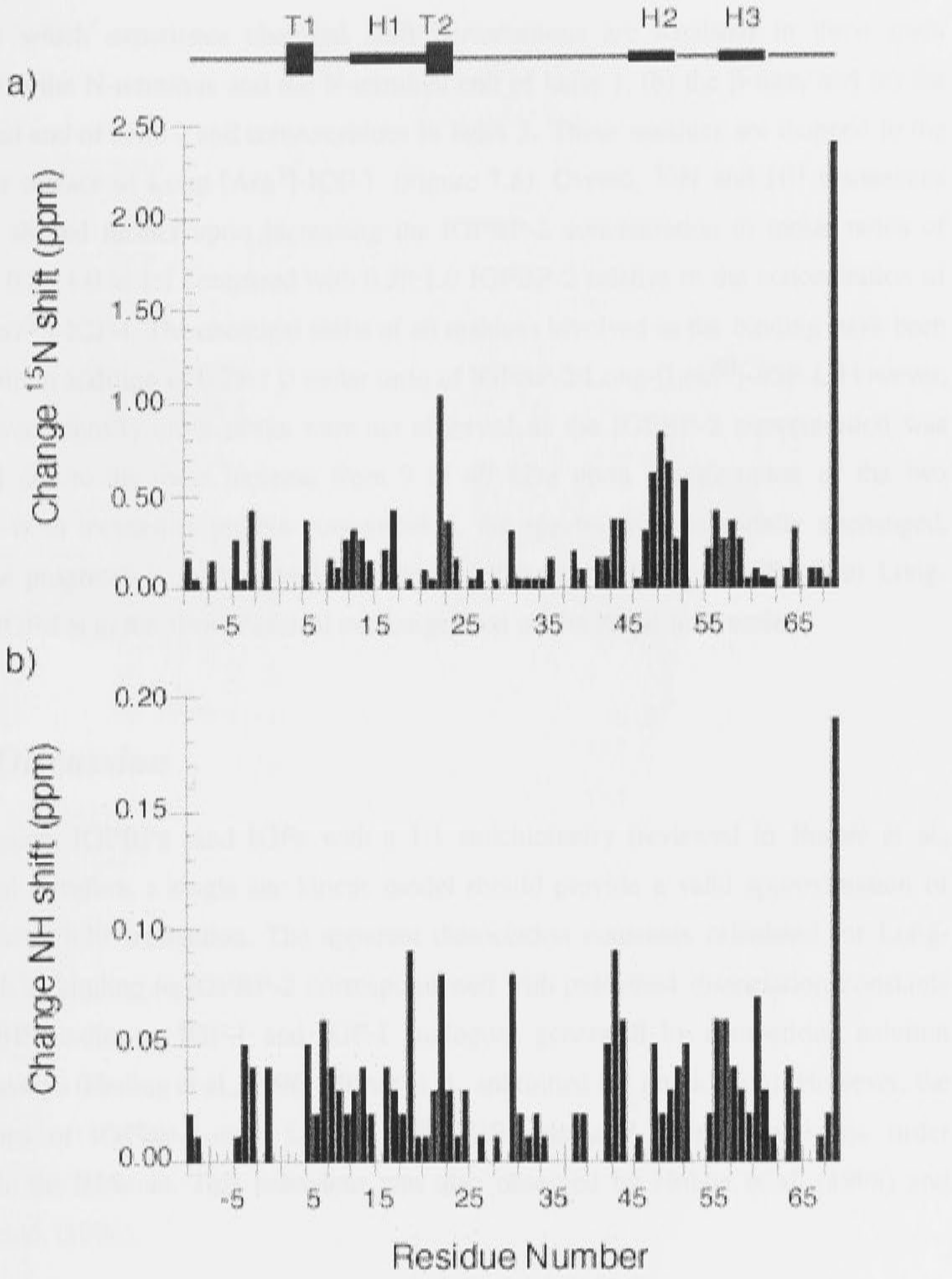

Figure 7.5: a) ^{15}N chemical shift changes in Long-[Leu^{60}]-IGF-I induced upon binding of IGFBP-2 to Long-[Leu^{60}]-IGF-I as a function of residue number at a molar ratio of 0.28:1.0 IGFBP-2:Long-[Leu^{60}]-IGF-I. b) Amide proton chemical shift changes in Long-[Leu^{60}]-IGF-I induced upon binding of IGFBP-2 to Long-[Leu^{60}]-IGF-I as a function of residue number at a molar ratio of 0.28:1.0 IGFBP-2:Long-[Leu^{60}]-IGF-I.

Residues which experience chemical shift perturbations are localized in three main regions; (a) the N-terminus and the N-terminal end of helix 1, (b) the β-turn, and (c) the C-terminal end of helix 2 and some residues in helix 3. These residues are mapped to the molecular surface of Long-[Arg^3]-IGF-I. (Figure 7.6). Overall, ^{15}N and H^N resonances were not shifted further upon increasing the IGFBP-2 concentration to molar ratios of 0.48:1.0, 0.83:1.0 or 1:1 compared with 0.28:1.0 IGFBP-2 relative to the concentration of Long-[Leu^{60}]-IGF-I. The chemical shifts of all residues involved in the binding have been affected upon addition of 0.28:1.0 molar ratio of IGFBP-2:Long-[Leu^{60}]-IGF-I. However, many lower intensity cross-peaks were not observed as the IGFBP-2 concentration was increased due to the mass increase from 9 to 40 kDa upon complexation of the two proteins. With increasing protein concentration, the spectra were essentially unchanged, other than progressive line broadening, indicating that the binding of IGFBP-2 to Long-[Leu^{60}]-IGF-I is in the slow chemical exchange limit on the NMR time scale.

7.4 Discussion

In circulation, IGFBPs bind IGFs with a 1:1 stoichiometry (reviewed in Baxter et al., 1994) and therefore, a single site kinetic model should provide a valid approximation of the IGFBP-2/IGF interaction. The apparent dissociation constants calculated for Long-[Leu^{60}]-IGF binding to IGFBP-2 correspond well with published dissociation constants for IGFBP binding to IGF-I and IGF-I analogues generated by competition solution binding assays (Heding et al., 1996; Milner et al., submitted for publication). However, the interactions of IGFBP-2 with Long-[Leu^{60}]-IGF-I deviated from pseudo-first order kinetics in the BIAcore. This behaviour was also observed by Hobba et al. (1998) and Heding et al. (1996).

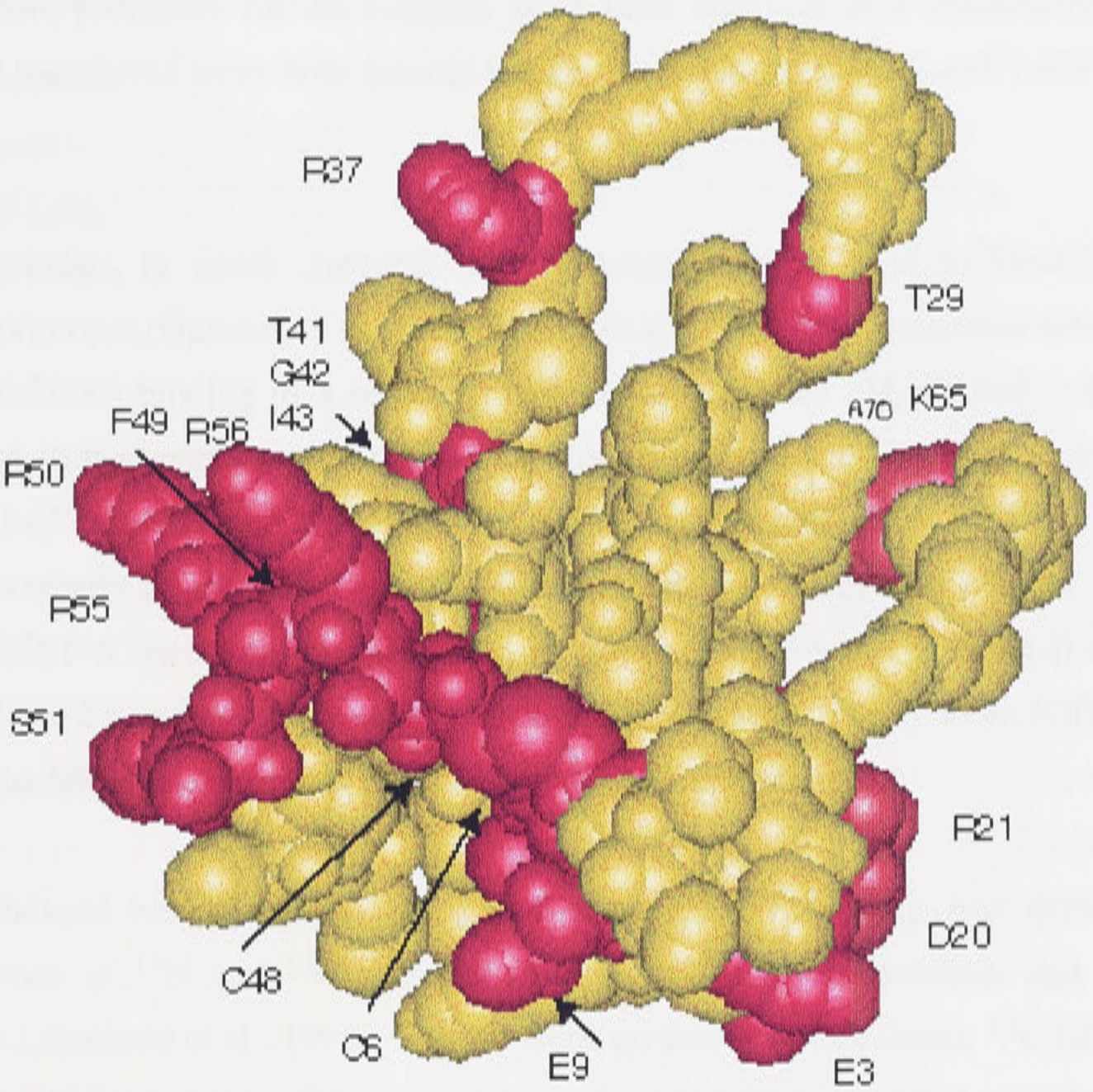

Figure 7.6: Molecular surface of Long-[Arg^3]-IGF-I with the residues of Long-[Leu^{60}]-IGF-I for which chemical shift perturbations are detected upon complexation with IGFBP-2 at 30^{o}C shown in red.

Jansson et al. (1998) formed the IGF-I/IGFBP-1 complex at dilute concentrations with an excess of IGFBP-1, in 7 ml of 1 mM sodium citrate buffer at pH 3.4 and subsequently concentrated by reduction of the sample volume to minimise the previously noted aggregation tendencies of both IGF-I and IGFBP-1 when in free solution (Cooke et al., 1991; Sato et al., 1992; Jansson et al., 1998). Sodium acetate buffer was used to maintain the complex at pH 3.0. Sato et al. (1992; 1993) reduced the self association of IGF-I at 2-10 mM protein concentration and improved the spectral quality by the addition of 10% or 20% acetic acid. In the study by Jansson et al. (1998) and the present study, acetic acid was not used as it affects the binding affinity of IGFs for IGFBPs (Jansson et al., 1998). In the work reported herein, no precipitation occurred on formation of the Long-[Leu^{60}]-IGF-I/IGFBP-2 complex. Titrating IGFBP-2 into the NMR sample containing Long-[Leu^{60}]-IGF-I to form the Long-[Leu^{60}]-IGF-I/IGFBP-2 complex simplified the

assignment procedure for the complex since most chemical shift assignments could be directly transferred from those determined for uncomplexed Long-[Leu^{60}]-IGF-I (Chapter 3).

Recent studies, in which chemical shift perturbations were used to identify structural binding epitopes (Grzesiek et al., 1996; Craven et al., 1996) have involved a small molecule (M_r 500-2000) binding to a relatively large labeled protein (M_r 20,000 - 25,000), i.e. chemical shift changes were observed on the larger molecule. Recently, mini-IGFBP-5 (Ala^{40}-Ile^{92} of the N-terminal domain of IGFBP-5) was complexed to IGF-I and IGF-II by successively adding IGF-I and IGF-II and NMR resonances of the larger ^{15}N-labeled mini-IGFBP-5 were measured (Kalus et al., 1998). This work with Long-[Leu^{60}]-IGF-I and IGFBP-2 presents the successive docking of the IGFBP family to an IGF-I analogue where the NMR resonances of the IGF-I analogue were measured.

The structural binding epitope of IGF-I for IGFBP-1 binding was defined by the comparison of ^{15}N and ^{13}C chemical shifts of IGF-I in the presence and absence of IGFBP-1 (Jansson et al., 1998). For the work presented in this thesis, ^{13}C-labeled Long-[Leu^{60}]-IGF-I was not available for studying the interaction between Long-[Leu^{60}]-IGF-I and IGFBP-2, thus, only ^{15}N chemical shifts of Long-[Leu^{60}]-IGF-I in the absence and presence of IGFBP-2 were used to define a structural binding epitope of Long-[Leu^{60}]-IGF-I. The lack of ^{13}C-labeled material is unfortunate since four of the five prolines in IGF-I were determined to have chemical shifts affected by IGFBP-1 binding (Jansson et al., 1998) and any chemical shift perturbations of these residues in Long-[Leu^{60}]-IGF-I cannot be observed. It should be noted that Jansson et al. (1998) encountered difficulties in making assignments for free IGF-I at 30 ^{o}C and that additional, but not all, IGF-I resonances could be assigned unambiguously in the complex with IGFBP-1, despite the increase in monomer molecular mass from 7 to 30 kDa. Assignments of Long-[Leu^{60}]-IGF-I could be made with similar levels of confidence for the protein in both the free and complexed state at 30 ^{o}C.

The chemical shift differences found in Long-[Leu^{60}]-IGF-I when comparing the molecule alone and in complex with IGFBP-2 are mapped onto the Long-[Arg^{3}]-IGF-I structure in Figure 7.6 since the three-dimensional solution structure of Long-[Leu^{60}]-IGF-I was not determined. Long-[Arg^{3}]-IGF-I is a valid model for mapping the perturbed residues of Long-[Leu^{60}]-IGF-I since the pattern of secondary structure observed in

Long-[Leu^{60}]-IGF-I was almost identical to that of Long-[Arg^{3}]-IGF-I and many of the same long-range NOEs were observed (Chapter 4). For comparison, the chemical shift differences found in IGF-I between the free and IGFBP-1 complexed forms (Jansson et al., 1998) are mapped onto the IGF-I structure (Cooke et al., 1991) in Figure 7.7. The IGF-I resonances perturbed by IGFBP-1 binding include Pro^{2}, Glu^{3}, Cys^{6}, Gly^{7}, Pro^{28}, Thr^{29}, Gly^{30}, Gly^{32}, Arg^{36}, Arg^{37}, Gln^{40}, Thr^{41}, Gly^{42}, Ser^{51}, Pro^{63}, Lys^{65}, Pro^{66}, Lys^{68} and Ala^{70}. The perturbed resonances of Long-[Leu^{60}]-IGF-I on IGFBP-2 binding which are also perturbed resonances of IGF-I upon IGFBP-1 binding are Glu^{3}, Cys^{6}, Thr^{29}, Arg^{37}, Gln^{40}, Thr^{41}, Gly^{42}, Ser^{51}, Lys^{65} and Ala^{70} (Figure 7.5). Interestingly, no residues in helix 2 of IGF-I (Ile^{43} to Arg^{50}) experienced chemical shift perturbations upon IGFBP-1 binding, while six residues of α-helix 2 of Long-[Leu^{60}]-IGF-I experienced chemical shift changes. This difference between IGF-I and Long-[Leu^{60}]-IGF-I upon IGFBP binding may reflect a different conformation of helix 2 between the two proteins due to the Tyr^{60}Leu mutation, or subtle differences between how IGFBP-1 and IGFBP-2 interact with IGFs, or even the difficulties Jansson et al. (1998) encountered in assigning this segment of uncomplexed IGF-I.

The amino acid residues Glu^{3}, Phe^{49}, Arg^{50} and Ser^{51} of IGF-I were identified by Dubaquié and Lowman (1999) in alanine screening studies as affecting IGFBP-1 affinity. These residues are a small portion of the interaction surface identified by the present NMR study, with some residues outside this surface. The structural binding epitope encompasses those residues whose structural orientation is altered by the IGFBP interaction, while the functional binding epitope encompasses those residues which are directly involved in binding to sites on the IGFBPs. It is likely that the functional binding epitope is significantly smaller than the structural binding epitope in line with findings from IGF-I/IGFBP-1 interaction (Jansson et al., 1998) and the interaction of human growth hormone with the soluble form of its receptor (Clackson et al., 1995).

An interesting observation in the present study is that the chemical shifts of some residues in the N-terminal extension and the C-terminus of Long-[Leu^{60}]-IGF-I are perturbed by the association of IGFBP-2 (Figures 7.5 and 7.6). Jansson et al. (1998) also noted that residues Lys^{68}-Ala^{70} of IGF-I were perturbed by the association of IGFBP-1, and the authors attributed this as being due to two alternative conformations adopted by the C-terminus of IGF-I, one of which is "frozen out" upon complexation. The chemical shift changes in the residues in the N-terminal extension and the C-terminus of Long-[Leu^{60}]-

IGF-I on IGFBP-2 binding may result from a different conformational population or indirect structural changes, rather than direct contact with the binding protein. There are two proline residues in both the N-terminal extension and the C-terminus, which suggests the possibility of altered conformations arising from *cis-trans* isomerism of these four prolines. In order to make a detailed comparison of the effect the N-terminal extension of Long-[Leu^{60}]-IGF-I has on the IGFBP-2 binding affinity it would be necessary also to investigate the chemical shift perturbations resulting from IGFBP-2 complexation with either [Leu^{60}]-IGF-I or IGF-I. Asp^{20} of Long-[Leu^{60}]-IGF-I also experiences a large ^{15}N chemical shift change on IGFBP-2 complexation along with smaller perturbations experienced by Arg^{21} and Gly^{22}, which comprise the β-turn at the end of helix 1. It is possible that these chemical shift changes are also the result of an altered conformation of the turn, however direct contact with the binding protein cannot be ruled out. This may also partially explain the ambiguity associated with categorising the type of β-turn in this region, which was discussed in Chapter 4.

Jansson et al. (1998) could not assign the disulfide-connected residues, Cys^{6} and Cys^{48} in free IGF-I, however, cross peaks were observed for Cys^{6} after the addition of IGFBP-1. The authors attributed this to a conformational exchange between the *cis* and *trans* isoforms of the disulfide bond upon IGFBP-1 binding. The authors propose that this exchange between the *cis* and *trans* isoforms may be responsible for the two observed C-terminal conformations. In contrast, both Cys^{6} and Cys^{48} could be assigned in free Long-[Leu^{60}]-IGF-I and the intensity of the cross peaks in the 2D ^{15}N HSQC spectrum was not significantly altered upon the addition of 25% IGFBP-2. In support of the proposed *cis-trans* isomerism of IGF-I (Jansson et al., 1998), both Cys^{6} and Cys^{48} in Long-[Leu^{60}]-IGF-I experience ^{15}N chemical shift perturbations upon the addition of IGFBP-2 with the shift change experienced by Cys^{48} being quite substantial at 0.84 ppm.

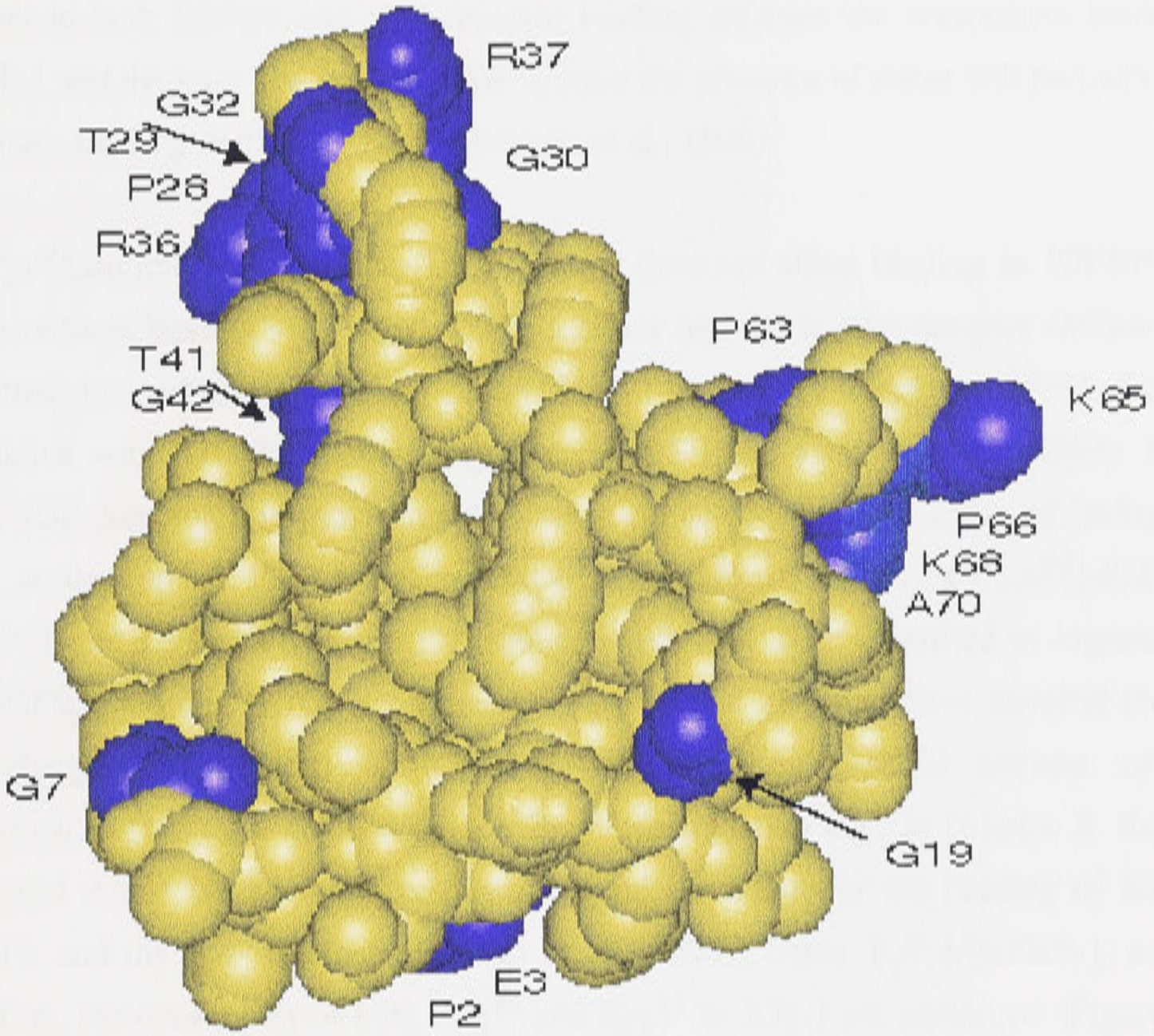

Figure 7.7: Molecular surface of IGF-I (Cooke et al., 1991) showing the residues for which chemical shift perturbations were detected upon complexation with IGFBP-1 (Jansson et al., 1998) in blue. IGF-I is rotated by approximately 90° about the z-axis relative to the orientation of Long-[Arg^3]-IGF-I in Figure 7.6.

Arginine sidechain guanidino groups are often found to be important in stabilising interactions between proteins (Pascal et al., 1994; Jansson et al., 1998). Jansson et al. (1998) identified Arg^{50}, Arg^{36} and Arg^{37} of IGF-I, by mutation of these arginine residues to alanines, as being functionally involved with IGFBP-1 binding, while Arg^{21}, Arg^{55} and Arg^{56} were found not to be functionally involved in binding. No mutations of arginine residues were explored in the present study. However, the backbone ^{15}N resonance of Arg^{50} of Long-[Leu^{60}]-IGF-I was perturbed by the addition of IGFBP-2 (Figure 7.5), suggesting a stabilising role for the Arg^{50} sidechain upon IGFBP-2 complex formation. The IGF-I arginines are highly conserved in vertebrates with the only common variant being an arginine to glutamine substitution at position of 50 in chicken, frog and salmon suggesting important functions for these residues. Arginine residues 36, 37, 21 and 56 are also involved with binding to the type 1 IGF receptor. The importance of these arginine

residues in both IGFBP and IGF receptor binding explains the competitive binding of IGFBP-1 and the type 1 receptor for IGF-I, since the presence of either will partially block the other's binding surface on IGF-I (Jansson et al., 1998).

The $Tyr^{60}Leu$ mutation of Long-[Leu^{60}]-IGF-I does not affect binding to IGFBP-2 but greatly reduces binding to the type 1 IGF receptor and the insulin receptor (Milner et al., submitted for publication). Long-[Leu^{60}]-IGF-I selectively displaces IGF-I from its association with IGFBPs, since Long-[Leu^{60}]-IGF-I has a much lower affinity for the type 1 IGF receptor, and thus promotes biological activity through action of endogenous IGF-I at the type 1 IGF receptor. Although the affinity of Long-[Leu^{60}]-IGF-I for IGFBPs is not significantly affected, residues which have been identified as important for both IGFBP and type 1 IGF receptor binding of IGF-I may not show identical chemical shift changes on binding of IGFBP-2 to Long-[Leu^{60}]-IGF-I because of local conformational changes induced by the $Tyr^{60}Leu$ mutation e.g., in α-helix 2. Residues Arg^{36} and Arg^{37} are two such residues that are important for the binding of IGF-I to IGFBPs and the IGF receptor (Jansson et al., 1998). Upon IGF-I/IGFBP-1 complex formation, the resonances of both Arg^{36} and Arg^{37} of IGF-I are perturbed (Figure 7.7). However, upon complexation of IGFBP-2 with Long-[Leu^{60}]-IGF-I, the Arg^{37} resonance of Long-[Leu^{60}]-IGF-I was perturbed, but not the Arg^{36} resonance. This difference may be attributed to an altered conformation in this C-domain segment induced by the $Tyr^{60}Leu$ mutation (see section 4.5).

A structural binding epitope on Long-[Leu^{60}]-IGF-I for IGFBP-2 has been defined by chemical shift mapping studies. Most of the residues involved in the binding interaction are located on one face of the molecule, however, Long-[Leu^{60}]-IGF-I has quite broad specificity determinants for interaction with IGFBP-2.

CHAPTER 8

Conclusions

The work presented in this dissertation adds to the wealth of structural knowledge of IGFs with the NMR studies of two IGF-I analogues, Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I. In particular, the first examples of the analysis of backbone dynamics for any member of the IGF-I family are presented. IGF-I and IGF-II are important mitogenic agents, while Long-[Arg^3]-IGF-I shows increased mitogenic activity and Long-[Leu^{60}]-IGF-I promotes the release of endogenous IGF-I from its association with IGFBPs and, thereby indirectly, the action of IGF-I through the type 1 IGF receptor.

Improvements in the resolution of NMR spectroscopy have enabled high quality spectra to be obtained at relatively low protein concentrations. 2D ^{15}N-HSQC spectra were obtained with sample concentrations of Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I as low as 0.4 mM, while 3D ^{15}N-NOESY-HSQC and 3D ^{15}N-NOESY-HSQC and 3D ^{15}N-TOCSY-HSQC spectra were obtained with protein concentrations of between 0.5 and 1.0 mM. Homonuclear NMR studies of IGF-I have, in the past, been hindered by the tendency of IGF-I to aggregate in solution at concentrations of 3-10 mM and pH above 5 (Cooke et al., 1991; Sato et al., 1992). In the work presented in this thesis, the amount of protein necessary to obtain a high resolution structure was reduced, enabling cost effective production of ^{15}N-labeled protein, while the lower sample concentrations used reduced the prevalence of aggregation. Despite working with Long-[Arg^3]-IGF-I concentrations of around 1.0 mM, sedimentation equilibrium experiments revealed the presence of 20% dimer. The effects of dimerisation, however, were not manifest in NOESY spectra nor did they affect the interpretation of the backbone dynamics, since there is no evidence of intermonomer peaks.

A central feature of all IGF structures determined thus far is retention of the three α-helices that enclose a hydrophobic core. The three dimensional structure determination of Long-[Arg^3]-IGF-I and the secondary structure determination of Long-[Leu^{60}]-IGF-I supports the retention of the hydrophobic core of the IGF-I domain. Despite the addition of a 13 amino acid extension at the N-terminus of the protein, the approximate relative alignment of the three helices remains unaltered in comparison with IGF-I, with the exception of a slight reorientation of the helices. The major variations in the structure of Long-[Arg^3]-IGF-I, compared with IGF-I, occur at the N-terminus with a substantial reorientation of the N-terminal three residues of the IGF-I domain. The sidechain of the residue at position three is removed from its location between helices 1 and 3 in IGF-I to a highly exposed location in Long-[Arg^3]-IGF-I. Thus, the reduced binding of Long-

[Arg^3]-IGF-I to IGFBPs does not simply result from the electrostatic effect of the Glu^{3}Arg mutation, but also from a spatial effect caused by a conformational relocation of the relative positions of Arg^3 and Phe^{49} compared with the positions of Glu^3 and Phe^{49} in IGF-I, residues involved in binding to IGFBPs.

Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I undergo non-uniform and complicated backbone dynamics. The most ordered parts of Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I are the α-helices. Nevertheless, the average heteronuclear NOEs for Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I of 0.55±0.12 and 0.69±0.15 respectively, are significantly less than the value observed in well defined secondary structure of rigid proteins (>0.75). The reduced heteronuclear NOEs suggest that the helices are not completely rigid and this is supported by the relatively rapid amide proton exchange rates compared to the helices of other proteins. Despite having three well-defined α-helices, Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I are clearly more flexible than most proteins.

The residues of the N-terminal extension of both Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I have negative heteronuclear NOEs, with the exception of Phe^{-3} and Val^{-2}, suggesting a degree of disorder. However, the amide protons of several residues in the N-terminal extension of Long-[Arg^3]-IGF-I and Long-[Leu^{60}]-IGF-I exhibit slow exchange with the bulk solvent, which suggests that any dynamic disorder is correlated to protect the hydrophobic residues in the N-terminal extension from contact with the aqueous solvent. The C-domain loop (Gly^{30}-Gln^{40}) has heteronuclear NOEs indicative of substantial but restricted motions as might be expected in a long loop.

Quantitative analysis of the relaxation data according to the model-free approach is complicated by the presence of 20% dimer, since aggregation systematically increases T_1 and decreases T_2. However, analysis performed using an axially symmetric rotational diffusion tensor is sufficient to show that the overall correlation time is longer by approximately a factor of two than other proteins of similar size. It has been suggested that a broad specificity of protein interaction with ligands or receptors is correlated with a certain degree of flexibility (Carr et al., 1997). It may be necessary for IGF-I to be sufficiently flexible in order to interact with the insulin and the type 1 IGF receptors as well as a family of six binding proteins.

Specific binding determinants of Long-[Leu^{60}]-IGF-I for IGFBP-2 were probed by

acquiring 2D ^{15}N-HSQC spectra with and without IGFBP-2 present. Long-[Leu^{60}]-IGF-I residues involved in the binding interactions with IGFBP-2 were identified by a significant ^{15}N chemical shift change upon complex formation with IGFBP-2. A number of residues were perturbed upon complex formation, including Glu^{3}, Cys^{48}, Phe^{49}, Arg^{50} and Ser^{51}, residues previously identified as involved in IGFBP complex formation (Jansson et al., 1998; Dubaquié and Lowman, 1999). These residues are all located on one face of the protein that is composed of the N-terminus of IGF-I, the N-terminal end of α-helix 2 and the region between α-helices 2 and 3, with the sidechains exposed to solvent.

The choice of the Val^{-2}-Asn^{-1} linker between the remainder of the N-terminal extension derived from pGH and the IGF-I domain was purely fortuitous. The primary reasons for choosing this dipeptide were to introduce a location susceptible to cleavage by hydroxylamine at Asn^{-1}-Gly^{1}, while Val^{-2}-Asn^{-1} provides a restriction enzyme cleavage site in the DNA construct.. However, Asn has a preference for being in the first position in turn structures, Gly has a preference for being in turns and Pro has a preference for being at position 3 in turns. This introduces the unique structural element of a turn incorporating residues Asn^{-1}, Gly^{1}, Pro^{2} and Arg^{3}. Research is in progress to find alternative linker sequences which can be cleaved proteolytically to give IGF-I or IGF-II or analogues of these proteins without the "long" N-terminal extension (G.L. Francis and S. Lien, personal communication). If an alternative linker is used with the fusion peptide for production of an IGF analogue with the extension intact, the turn is not likely to be maintained unless residues are used in the linker sequence which also have a propensity to form turn structures. The turn between Asn^{-1} and Arg^{3} in Long-[Arg^{3}]-IGF-I is a crucial component in altering the affinity of Long-[Arg^{3}]-IGF-I for IGFBPs and the receptors, since it brings the extension close to the IGF domain and alters the sidechain conformation of residue 3.

NMR spectroscopy will prove an important tool for future IGF research. Rapid growth has taken place in the field of NMR spectroscopy in the last few years and its application to the study of macromolecules. The present work has provided an important structural basis for future work. The structure determination of the family of IGFBPs, crucial for understanding their interactions with IGFs, is now feasible. NMR methods will, in the future, be able to identify residues of IGFs, IGFBPs or IGF receptors involved in the structural and functional binding sites. The flexibility of the IGFs and the conformational changes undergone to accommodate an approaching ligand can be investigated using backbone dynamics. The mobility of the IGFs is perhaps more significant than site

directed mutations in defining the specificity of interactions. The availability of $^{13}C/^{15}N$ double labeled samples would result in structures of higher precision enabling more detailed comparison of the orientation of key sidechains and improved backbone dynamics analyses.

REFERENCES

References

Abdel-Meguid, S.S., Shei, H-S., Smith, W.W., Dayringer, H.E., Violand, B.N. and Bentle, L.A. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6434-6437

Abragam, A. (1961) *Principles of Nuclear Magnetism*, Oxford University Press, London.

Adams, S.O., Nissley, S.P., Handwerger, S. and Rechler, M.M. (1983) *Nature* **302**, 150-153

Allerhand, A., Doddrell, D. and Komoroski, R. (1971) *J. Chem. Phys.* **55**, 189-198

Andress, D.L., Loop, S.M., Zapf, J. and Kiefer, M.C. (1993) *Biochem. Biophys. Res. Commun.* **195**, 25-30

Anfinsen, C.B. (1973) *Science* **181**, 223-230

Arai, T., Clarke, J., Parker, A., Busby, W. Jr., Nam, T. and Clemmons, D.R. (1996) *J. Biol. Chem.* **271**, 6099-6106

Archer, S.J., Ikura, M., Torchia, D.A. and Bax, A. (1991) *J. Magn. Reson.* **95**, 636-641

Aue, W.P., Bartholdi, E. and Ernst, R.R. (1976) *J. Chem. Phys.* **64**, 2229-2246

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds) (1998) in: *Current Protocols in Molecular Biology*, John Wiley, New York

Bach, L.A., Hsieh, S., Sakano, K., Fujiwara, H., Perdue, J.F. and Rechler, M.M. (1993) *J. Biol. Chem.* **268**, 9246-9254

Bach, L.A. and Rechler, M.M. (1995) *Diabetes Review* **3**, 38-61

Bagley, C.J., May, B.L., Szabo, L., McNamara, P.J., Ross, M., Francis, G.L., Ballard, F.J. and Wallace, J.C. (1989) *Biochem. J.* **259**, 665-671.

Baker, D., Sohl, J.L. and Agard, D.A. (1992) *Nature* **356**, 263-265

Ballard, F.J., Francis, G.L., Ross, M., Bagley, C.J., May, B. and Wallace, J.C. (1987) *Biochem. Biophys. Res. Commun.* **149**, 398-404

Ballard, F.J., Ross, M., Upton, F.M. and Francis, G.L. (1988) *Biochem. J.* **249**, 721-726

Ballard, F.J., Walton, P.W., Dunshea, F.R., Francis, G.L. and Tomas, F.M. (1994) in *The insulin-like growth factors and their regulatory proteins* (Baxter, R.C., Gluckman, P.D. and Rosenfeld, R.G. eds) pp. 131-138, Excerpta Medical, New York

Bartels, C., Xia, T-H., Billeter, M., Güntert, P. and Wüthrich, K. (1995) *J. Biomol. NMR* **6**, 1-10

Basus, V.J. (1989) *Methods Enzymol.* **177**, 132-149

Baum, J., Dobson, C.M., Evans, P.A. and Hanley, C. (1989) *Biochemistry* **28**, 7-13

Bautista, C.M., Baylink, D.J. and Mohan, S. (1991) *Biochem. Biophys. Res. Commun.* **176**, 756-763

Baxter, R.C. (1994) *Horm. Res. (Basel)* **42**, 140-144

Baxter, R.C. and Martin, J.L. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6898-6902

Bayne, M.L., Applebaum, J., Chicchi, G.G., Hayes, N.S., Green, B.G. and Cascieri, M.A. (1988) *J. Biol. Chem.* **263**, 6233-6239

Bayne, M.L., Applebaum, J., Chicchi, G.G., Miller, R.E. and Cascieri, M.A. (1990) *J. Biol. Chem.* **265**, 15648-15652

Bayne, M.L., Applebaum, J., Underwood, D., Chicchi, G.G., Green, B.G., Hayes, N.S. and Cascieri, M.A. (1989) *J. Biol. Chem.* **264**, 11004-11008

Bayne, M.L., Cascieri, M.A., Kelder, B., Appelbaum, J., Chicchi, G.G., Shapiro, J.A., Pasleau, F. and Kopchick, J.J. (1987) *Proc. Natl. Acad. Sci. USA* **85**, 2638-2642.

Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.E. Jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1997) *J. Mol. Biol.* **112**, 535

Bloch, F., Hansen, W.W. and Packard, M.E. (1946) *Phys. Rev.* **69**, 127

Blundell, T.L., Bedarkar, S., Rinderknecht, E. and Humbel, R.E. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 180-184

Blundell, T.L., Bedarkar, S. and Humbel, R.E (1983) *Fed. Proc.* **42**, 2542-2547

Bodenhausen, G. and Ruben, D. (1980) *Chem. Phys. Lett.* **69**, 185-189

Broadhurst, R.W., Hardman, C.H., Thomas, J.O. and Laue, E.D. (1995) *Biochemistry* **34**,

16608-16617

Brüschweiler, R., Liao, X. and Wright, P.E. (1995) *Science* **268**, 886-669

Buck, M., Schwalbe, H. and Dobson, C.M. (1995) *Biochemistry* **34**, 13219-13232

Burum, D.P. and Ernst, R.R. (1980) *J. Magn. Reson.* **39**, 163

Bystrov, V.F. (1976) *Prog. NMR Spectroscopy* **10**, 41-81

Cai, M., Huang, Y., Sakaguchi, K., Clore, G.M., Gronenborn, A.M. and Craigie, R. (1998) *J. Biomol. NMR* **11**, 97-102

Campbell, I.D. (1991) *Biochem. Soc. Trans.* **19**, 243-248

Campbell, P.G. and Andress, D.L. (1997) *Am. J. Physiol. Endocrinol. Metab.* **36**, 1005-1013

Canalis, E., Centrella, M., Burch, W. and McCarthy, T.L. (1989) *J. Clin. Invest.* **83**, 60-65

Cara, J.F., Mirmira, R.G., Nakagawa, S.H. and Tager, H.S. (1990) *J. Biol. Chem.* **265**, 17820-17825

Carr, H.Y. and Purcell, E.M. (1954) *Phys. Rev.* **4**, 630-638

Carr, P.A., Erickson, H.P. and Palmer, A.G., III (1997) *Structure* **5**, 949-959

Cascieri, M.A. and Bayne, M.L. (1989) in: *Molecular and cellular biology insulin-like growth factors and their receptors*, LeRoith, D. and Raizada, M.K. eds. Plenum, New York. p. 225

Cascieri, M.A., Chicchi, G.C., Applebaum, J., Hayes, N.S., Green, B.C. and Bayne, M.L (1988) *Biochemistry* **27**, 3228-3233

Cascieri, M.A., Chicchi, G.G., Applebaum, J., Green, B.G., Hayes, N.S. and Bayne, M.L. (1989) *J. Biol. Chem.* **264**, 2199-2202

Cavanagh, J., Fairbrother, W.J., Palmer, A.G. III, and Skelton, N.J. (1996) *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, New York

Cheng, J-W., Lepre, C.A., Chambers, S.P., Fulghum, J.R., Thomson, J.A. and Moore, J.M. (1993) *Biochemistry* **32**, 9000-9010

Chernausek, S.D., Smith, C.E., Duffin, K.L., Busby, W.H., Wright, G. and Clemmons, D.R. (1995) *J. Biol. Chem.* **270**, 11377-11382

Clackson, T. and Wells, J.A. (1995) *Science* **267**, 383-386

Clemmons, D.R. (1997) *Cytokine Growth Factor Review* **8**, 45-62

Clemmons, D.R:, Dehoff, M.L., Busby, W.H., Bayne, M.L. and Cascieri, M.A. (1998) *Endocrinol.* **131**, 890-95

Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C. and Gronenborn, A.M. (1990a) *J. Am. Chem. Soc.* **112**, 4989-4991

Clore, G.M., Driscoll, P.C., Wingfield, P.T. and Gronenborn, A.M. (1990b) *Biochemistry* **29**, 7387-7401

Cohick, W.S. and Clemmons, D.R. (1993) *Ann. Rev. Physiol.* **55**, 131-153

Conover, C.A. (1995) *Prog. Growth Factor Res.* **6**, 301-309

Conover, C.A. and Powell, D.R. (1991) *Endocrinology* **129**, 710-716

Cooke, R.M., Harvey, T.S. and Campbell, I.D. (1991) *Biochemistry* **30**, 5484-5491

Craven, J.C., Whitehead, B., Jones, S.K.A., Thulin, E., Blackburn, M. and Waltho, J.P. (1996) *Biochemistry* **35**, 10287-10299

Daughaday, W.H., Hall, K., Raben, M.S., Salmon, W.D. Jr, van de Brandle, J.L. and van Wyk, J.J. (1972) *Nature* **235**, 107

Daughaday, W.H. and Rotwein, P. (1989) *Endocrine Rev.* **10**, 68-91

DeChiara, T.M., Efstratiadis, A. and Robertson, E.J. (1990) *Nature* **345**, 78-80

DeLorimer, R. Hellinga, H.W. and Spicer, L.D. (1996) *Protein Science*, **5**, 2552-2565

De Vroede, M.A., Rechler, M.M., Nissley, S.P., Joshi, S., Burke, G.T. and Katsoyannis, P.G. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3010-3014

De Wolf, E., Gill, R., Geddes, S., Pitts, J., Wollmer, A. and Grötzinger, J. (1996) *Protein Sci.* **5**, 2193-2202

Dubaquié, Y. and Lowman, H.B. (1999) *Biochemistry* **38**, 6386-6396

Durham, S.K., Mohan, S., Liu, F., Baker, B.K., Lee, P.D., Hintz, R.L., Conover, C.A. and Powell, D.R. (1997) *Pediatric Res.* **42**, 335-341

Englander, S.W. (1975) *Ann. N.Y. Acad. Sci.* **244**, 10-27

Englander, S.W. and Kallenbach, N.R. (1984) *Q. Rev. Biophys.* **16**, 521-655

Ernst, M., Heath, J.K. and Rodan, G. (1989) *Endocrinology* **125**, 825-833

Ernst, R.R. and Anderson, W.A. (1966) *Rev. Sci. Inst.* **37**, 93-102

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Schoelson, S.E., Pawson, T., Forman-Kay, J.D. and Kay, L.E. (1994) *Biochemistry* **33**, 5984-6003

Farrow, N.A., Zhang, O., Forman-Kay, J.D. and Kay, L.E. (1997) *Biochemistry* **36**, 2390-2402

Feng, Y., Klein, B.K. and McWherter, C.A. (1996) *J. Mol. Biol.* **259**, 524-541

Fersht, A.R. and Serrano, L. (1993) *Curr. Opin. Struct. Biol.* **3**, 75-83

Forbes, B.E., Turner, D., Hodge, S.J., NcNeil, K.A., Forsberg, G. and Wallace, J.C. (1988) *J. Biol. Chem.* **273**, 4647-4652

Fowlkes, J. and Freemark, M. (1992) *Endocrinology* **131**, 2071-2076

Fowlkes, J.L., Serra, D.M., Rosenberg, C.K. and Thrailkill, K.M. (1995) *J. Biol. Chem.* **270**, 27481-27488

Francis, G.L., Read, L.C., Ballard, F.J., Ballard, F.J., Bagley, C.J., Upton, F.M., Gravestock, P.M. and Wallace, J.C. (1986) *Biochem. J.* **251**, 95-103

Francis, G.L., Ross, M., Ballard, F.J., Milner, S.J., Senn, C., McNeil, K.A., Wallace, J.C., King, R. and Wells, J.R.E. (1992) *J. Mol. Endocrinol.* **8**, 213-223

Francis, G.L., Upton, F.M., Ballard, F.J., McNeil, K.A. and Wallace, J.C. (1988) *Biochem. J.* **251**, 95-104

Froesch, E.R., Burgi, H., Ramsier, E.B., Bally, P. and Labhart, A. (1963) *J. Clin. Invest.* **42**, 1816-1834

Fägerstam, L.G., Frostell-Karlsson, A., Persson, B. and Ronnberg, I. (1992) *J.*

Chromatography **597**, 397-410

Gething, H.F. and Sambrook, J. (1992) *Nature* **335**, 33-45

Gill, R., De Meyts, P., De Wolf, E., Geddes, S., Grötzinger, J., HaBiepen, U., Murray-Rust, J., Pilts, J., Urso, B., Verma, C., Wallach, B., Wollmer, A. and Woods, S. (1996) *Protein Eng.* **9**, 1011-1019

Gill, S.C. and von Hippel, P.H. (1989) *Anal. Biochem.* **182**, 319-326

Gopinath, R., Walton, P.E. and Etherton, T.D. (1989) *J. Endocrinology* **120**, 231-236

Grzesiek, S. and Bax, A. (1993) *J. Am. Chem. Soc.* **115**, 12593-12594

Grzesiek, S., Bax, A., Hu, J-S., Kaufman, J., Palmer, I., Stahl, S.J., Tjandra, N. and Winglfield, P.T. (1997) *Protein Sci.* **6**, 1248-1263

Grzesiek, S., Stahl, S.J., Wingfield, P.T. and Bax, A. (1996) *Biochemistry* **35**, 10256-10261

Guler, H-P., Schmid, C., Zapf, J. and Froesch, E.R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2868-2872

Gutowsky H.S., McCall, D.W. and Schlincter, C.P. (1951) *Phys. Rev.* **84**, 589-590

Güntert, P., Braun, W., Billeter, M. and Wüthrich, K. (1989) *J. Am. Chem. Soc.* **111**, 3997-4004

Güntert, P., Mumenthalar, C. and Wüthrich, K. (1997) *J. Mol. Biol.* **273**, 283-298

Hall, K. and Sara, V.R. (1983) *Vitam. Horm.* **40**, 175-233.

Han, V.K.M., D'Ercole, A.J. and Lund, P.K. (1987) *Science* **236**, 193-197

Hari, J., Pierce, S.B., Morgan, D.O., Sara, V., Smith, M.C. and Roth, R.A. (1987) *EMBO J.* **6**, 3367-3371

Hashimoto, R., Ono, M., Fujiwara, H., Higashihashi, N., Yoshida, M., Enjohkimura, T. and Sakao, K. (1997) *J. Biol. Chem.* **272**, 27936-27942

Heding, A., Gill, R., Ogawa, Y., De Meyts, P., Shymko, R.M. (1996) *J. Biol. Chem.* **271**, 13948-13952

Higo, H., Duan, C., Clemmons, D.R. and Herman, B. (1997) *Biochem. Biophys. Res. Commun.* **239**, 706-709

Hintz, R.L. and Liu, F. (1977) *J. Clin. Endocrinol. Metab.* **45**, 988-995

Hiyama, Y., Niu, C., Silverton, J.V., Bavoso, A. and Torchia, D. A. (1988) *J. Am. Chem. Soc.* **110**, 2378-2383

Hobba, G.D., Forbes, B.E., Parkinson, E.J., Francis, G.L. and Wallace, J.C. (1996) *J. Biol. Chem.* **271**, 30529-30536

Hobba, G.D., Löthgren, A., Holmberg, E., Forbes, B.E., Francis, G.L. and Wallace, J.C. (1998) *J. Biol. Chem.* **273**, 19691-19698

Hober, S., Forsberg, G., Palm, G., Hartmanis, M. and Nilsson, R. (1992) *Biochemistry* **31**, 1749-1756

Hober, S., Hansson, A., Uhlén, M. and Nilsson, B. (1994) *Biochemistry* **33**, 6758-6761

Hua, Q.X., Narhi, L., Jia, W.H., Arakawa, T., Rosenfeld, R., Hawkins, N., Miller, J.A. and Weiss, M.A. (1996) *J. Mol. Biol.* **259**, 297-313

Huang, K., Flanagan, J.M. and Prestegard, J.H. (1999) *Protein Sci.* **8**, 203-241

Humbel, R.E. (1990) *Eur. J. Biochem.* **190**, 445-462

Hurd, R.E. and John, B.K. (1991) *J. Magn. Reson.* **92**, 658-668

Hynes, M.A., Van Wyk, J.J., Brooks, P.J., D'Ercole, A.J., Jansen, M. and Lund, P.K. (1987) *Mol. Endocrinol.* **1**, 233-242

Hyre, D.E. and Klevit, R.E. (1998) *J. Mol. Biol.* **279**, 929-943

Ikura, M., Kay, L.E. and Bax, A. (1990) *Biochemistry* **29**, 4659-4667

Jaenicke, R. (1991) *Biochemistry* **30**, 3149-3161

Jansson, M., Li, Y-C., Jendeberg, L., Anderson, S., Montelione, G.T. and Nilsson, B. (1996) *J. Biomol. NMR* **7**, 131-141

Jansson, M., Anderson, G., Uhlén, Nilsson, B. and Kördel, J. (1998) *J. Biol. Chem.* **273**, 24701-24707

Johnson, T.R., Blossey, B.K., Denko, C.W. and Ilan, J. (1989) *Mol. Endocrinol.* **3**, 580-587

Jones, D.N.M., Searles, M.A., Shaw, G.L., Churchill, M.E.A., Ner, S.S., Keeler, J., Travers, A.A. and Neuhaus, D. (1994) *Structure* **2**, 609-627

Jones, J.I. and Clemmons, D.R. (1995) *Endocrine Rev.* **16**, 3-34

Jones, J.I., Gockerman, A., Busby, W.H. Jr., Camacho-Hubner, C. and Clemmons, D.R. (1993) *J. Cell Biol.* **121**, 679-687

Kalus, W., Zweckstetter, M., Renner, C., Sanchez, Y., Georgescu, J., Grol, M., Demuth, D., Schuacher, R., Dony, C., Lang, K. and Holak, T.A. (1998) *EMBO J.* **17**, 6558-6572

Kandror, K.V. and Pilch, P.F. (1996) *J. Biol. Chem.* **271**, 21703-21708

Karplus, M. (1959) *J. Phys. Chem.* **30**, 10-15

Kay, L.E. (1995) *Curr. Opin. Struct. Biol.* **5**, 674-681

Kay, L.E., Torchia, D.A. and Bax, A. (1989) *Biochemistry* **28**, 8972-8979

Kay, L.E. and Gardner, K.H. (1997) *Curr. Opin. Struct. Biol.* **7**, 722-731

Kay, L.E., Keifer, P. and Saarinen, T. (1992) *J. Am. Chem. Soc.* **114**, 10663-10665

Kay, L.E., Muhandiram, D.R., Farrow, N.A., Aubin, Y. and Forman-Kay, J.D. (1996) *Biochemistry* **35**, 361-368

Kiess, W., Hoeflich, A., Yang, Y., Kessler, U., Flyvbjerg, A and Barenton, B. (1994) *Current Directions in Insulin-Like Growth Factor Research*, ed. Le Roith, D. and Raozada, M.K., New York.

King, R., Wells, J.R.E., Krieg, P., Snoswell, M., Brazier, J., Bagley, C.J., Wallace, J.C., Ballard, F.J., Ross, M. and Francis, G.L. (1992) *J. Mol. Endocrinol.* **8**, 29-41.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) *J. Mol. Graphics* **14**, 51-55

Kraulis, P.J. (1994) *J. Mol. Biol.* **243**, 696-718

Krebs H., Schmid, F.X. and Jaenicke, R. (1983) *J. Mol. Biol.* **169**, 619-635

Kurtz, A., Zapf, J., Eckardt, K.U., Clemmons, G., Froesch, E.R. and Bauer, C. (1988)

Proc. Natl. Acad. Sci. USA **85**, 7825-7829

Kuwajima, K., Garvey, E.P., Finn, B.E., Matthews, C.R. and Sugai, S. (1991) *Biochemistry* **30**, 7693-7703

Ladbury, J.E., Wynn, R., Thompson, J.A. and Sturtevant, J.M. (1995) *Biochemistry* **34**, 2148-2152

Laskowski, R.A., Rullman, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) *J. Biomol. NMR* **8**, 477-486

Lee, P.D.K., Conover, C.A. and Powell, D.R. (1993) *Proc. Soc. Exp. Biol. Med.* **204**, 4-29

Lee, L.K., Rance, M., Chazin, W.J. and Palmer, A.G., III (1997) *J. Biomol. NMR* **9**, 287-298

Lee, W. (1994) *FEBS Lett.* **350**, 87-90

Leroith, D., King, G. and Flier, J.S. (1997) *N. Eng. J. Med.* **336**, 633-640

Levitt Katz, L.E., Rosenfeld, R.G. and Cohen, S.P. (1995) *The Endocrinologist* **5 (1)**, 36-44

Lipari, G. and Szabo, A. (1982a) *J. Am. Chem. Soc.* **104**, 4546-4559

Lipari, G. and Szabo, A. (1982b) *J. Am. Chem. Soc.* **104**, 4559-4570

Lobel, P., Dahms, N.M. and Kornfeld, S. (1988) *J. Biol. Chem.* **263**, 2563-2570

Ludvigsen, S., Andersen, K.V. and Poulsen, F.M. (1991) *J. Mol. Biol.* **217**, 731-736

Luthi, C., Roth, B.V. and Humbel, R.E. (1992) *Eur. J. Biochem.* **205**, 483-490

Löfås, S. and Johnsson, B. (1990) *J. Chem. Soc. Chem. Commun.* **21**, 1526-1528

McAlister, M.S.B., Mott, H.R., van der Merwe, P., Campbell, I.D., Davis, S.J. and Driscoll, P.C. (1996) *Biochemistry* **35**, 5982-5991

McCusker, R.H., Busby, W.H., Dehoff, M.H., Camacho-Hubner, C. and Clemmons, D.R. (1991) *Endocrinology* **129**, 939-951

McCusker, R.H., Camacho-Hubner, C., Bayne, M.L., Cascieri, M.A. and Clemmons, D.R. (1990) *J. Cell Physiol.* **144**, 244-253

Mackay, J.P., Shaw, G.L. and King, G.F. (1996) *Biochemistry* **35**, 4867-4877

Maly, P. and Lüthi, C. (1988) *J. Biol. Chem.* **263**, 7068-7072

Mandel, A.M., Akke, M. and Palmer, A.G., III (1995) *J. Mol. Biol.* **246**, 144-163

Marion, D., Kay, L.E., Sparks, S.W., Torchia, D.A. and Bax, A. (1989) *J. Am. Chem. Soc.* **111**, 1515-1517

Martin, J., Langer, T., Boteva, R., Schramel, A., Horwic, A.I. and Hartl, F-U. (1991) *Nature* **352**, 36-42

Massagué, J. and Czech, M.P. (1982) *J. Biol. Chem.* **257**, 5038-5045

Mathews, L.S., Norstedt, G. and Palmiter, R.D. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9343-9347

Meiboom, S. and Gill, D. (1958) *Rev. Sci. Instrum.* **29**, 688-691

Mellas, J., Gavin, J.R., II. and Hammerman, M.R. (1986) *J. Biol. Chem.* **261**, 14437-14442

Meng, H., Burleigh, B.D. and Kelly, G.M. (1989) *J. Chromatography* **443**, 183-192

Miller, J.A., Narhi, L.O., Hua, Q-X., Rosenfeld, R., Arakawa, T., Rohde, M., Prestrelski, S., Lauren, S., Stoney, K.S., Tsai, L. and Weiss, M.A. (1993) *Biochemistry* **32**, 5203-5213

Milner, S.J., Butler, I.P., Elliot, P.W., Grosvenor, S.E., Francis, G.L., Farah, J., Miller, M., Read, L.C. and Ballard, F.J. (submitted to "Growth Factors" for publication)

Milner, S.J., Carver, J.A., Ballard, G.L. and Francis, G.L. (1999) *Biotech. and Bioengin.* **62**, 693-703

Minniti, C.P., Kohn, E.C., Grubb, J.H., Sly, W.S., Oh, Y., Müller, H.L., Rosenfeld, R.G. and Helman, L.J. (1992) *J. Biol. Chem.* **267**, 9000-9004

Mohan, S., Najao, Y., Honda, Y., Landale, E., Leser, U., Dony, C., Lang, K. and Baylink, D.J. (1995) *J. Biol. Chem.* **270**, 20424-20431

Morris, G.A. and Freeman, R. (1979) *J. Am. Chem. Soc.* **101**, 760-762

Moss, J. A., Francis, G.L., Ross, M., Wallace, J.C. and Ballard, F.J. (1991) *J. Biol. Chem*

266, 909-914

Murayama, Y., Okamoto, T., Ogata, E., Asano, T., Iiri, T., Katada, T., Ui, M., Grubb, J.H., Sly, W.S. and Nishimoto, I. (1990) *J. Biol. Chem.* **265**, 17456-17462

Murphy, L.J., Murphy, L.C. and Freisen, H.G. (1987) *Mol. Endocrinol.* **1**, 445-450

Murray-Rust, J., McLeod, A.N., Blundell, T.L. and Wood, S.P. (1992) *Bioessays* **14**, 325-331.

Nagayama, K., Wüthrich, K. and Ernst, R.R (1977) *Biochem. Biophys. Res. Commun.* **90**, 305-311

Narhi, L.O., Hua, Q-X., Tsutomu, A., Fox, G.M., Tsai, L., Rosenfeld, R., Holst, P., Miller, J.A. and Weiss, M.A. (1993) *Biochemistry* **32**, 5214-5221

Nishimoto, I., Hata, Y., Ogata, E., Kojima, I. (1987) *J. Biol. Chem.* **262**, 12120-12126

Oh, Y., Nagalla, S.R., Yamanaka, Y., Kim, H-S., Wilson, E. and Rosenfeld, R.G. (1997) *J. Biol. Chem.* **271**, 30322-30325

Olenjniczak, E.T. and Eaton, H.L. (1990) *J. Magn. Reson*, **87**, 628-632

Omichinski, J.G., Clore, G.M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S.J. and Gronenborn, A.M. (1993) *Science* **261**, 438-446

Oschkinat, H., Cieslar, C., Gronenborn, A.M. and Clore, G.M. (1989) *J. Magn. Reson.* **81**, 212-216

Oschkinat, H., Müller, T. and Dieckmann, T. (1994) *Angew. Chem. Int. Ed. Engl.* **33**, 277-293

Overhauser, A.W. (1955) *Phys. Rev.* **92**, 411-415

Palmer, A.G., III, Rance, M. and Wright, P.E. (1991) *J. Am. Chem. Soc.* **113**, 4371-4380

Pardi, A., Billeter, M. and Wüthrich, K. (1984) *J. Mol. Biol.* **180**, 741-751

Pascal, S.M., Muhandiram, D.R., Yamazaki, T., Forman-Kay, J.D. and Kay, L.E. (1994) *J. Magn. Reson.* **103**, 197-201

Pierson, R.W. and Temin, H.M. (1972) *J. Cell Physiology* **79**, 319-330

Piotto, M., Saudek, V. and Sklenár, V. (1992) *J. Biomol. NMR* **2**, 661-665

Pollock, J.R., Swenson, R.P. and Stockman, B.J. (1996) *J. Biomol. NMR* **7**, 225-235

Proctor, W.G. and Yu, F.C. (1950) *Phys. Rev.* **77**, 717

Purcell, E.M., Torrey, H.C. and Pound, R.V. (1946) *Phys. Rev.* **69**, 37-38

Qin, X., Strong, D.D., Baylink, D.J. and Mohan, S.J. (1998) *J. Biol. Chem.* **273**, 23509-23516

Raschdorf, F., Dahinden, R., Maerki, W., Richter, W. and Merryweather, J. (1988) *Biomed. Environ. Mass Spectrom.* **16**, 3-8

Rechler, M.M. (1991) in *Insulin-Like Growth Factors: Molecular and Cellular Aspects* (ed. Le Roith, D.) pp 87-110, CRC Press, Boca Raton, FL

Rechler, M.M. (1993) *Vitam. Horm.* **47**, 1-14

Rechler, M.M. (1997) *Endocrinol.* **138**, 2645-2647

Redfield, C. (1993) Resonance assignment strategies for small proteins, In; *NMR of Macromolecules: A Practical Approach*, Oxford University Press

Richardson, J.S. and Richardson, D.C. (1988) *Science* **240**, 1648-1652

Rinderknecht, E. and Humbel, R.E. (1976) *Proc. Natl. Acad. Sci. USA* **12**, 4379-4381

Rinderknecht, E. and Humbel., R.E (1978) *J. Biol. Chem.* **253**, 2769-2776

Ritvos, O., Ranta, T., Jalkanen, J., Suikkari, A.M., Voutilainen, R., Bohn, H. and Rutanen, E.M. (1988) *Endocrinology* **122**, 2150-2157

Rosen, C.J. (1997) *Bone* **21**, 217-223

Ross, M., Francis, G.L., szabo, L., Wallace, J.C. and Ballard, F.J. (1989) *Biochem. J.* **258**, 267-272

Roth, R.A. (1988) *Science* **19**, 1269-1271

Roth, B.V., Bürgisser, D.M., Lüthi, C. and Humbel, R.E. (1991) *Biochem. Biophys. Res. Comm.* **181**, 907-914

Rutanen, E-M. and Pekonen, F. (1990) *Acta Endocrinol.* **123**, 7-13

Sakano, K., Enjoh, T., Numata, F., Fujiwara, H., Marumoto, Y., Higashihashi, N., Sato, Y., Perdue, J.F. and Fujita-Yamaguchi, Y. (1991) *J. Biol. Chem.* **266**, 20626-20635

Salmon, W.D. and Daughaday, W.H. (1957) *J. Lab. Clin. Med.* **49 (6)**, 825-836

Sara, V.R. and Hall, K. (1990) *Physiol. Rev.* **70 (3)**, 591-611

Sato, A., Nishimura, S., Ohkubo, T., Kyogoku, Y., Koyama, S., Kobayashi, M., Yasuda, T. and Kobayashi, Y. (1992) *J.Biochem. (Tokyo)* **111**, 529-536

Sato, A., Nishimura, S., Ohkubo, T., Kyogoku, Y., Koyama, S., Kobayashi, M., Yasuda, T. and Kobayashi, Y. (1993) *Int. J. Pept. Protein Res.* **41**, 433-440

Scheiwiller, E., Guler, H-P., Merryweather, J., Schandella, C., Maerki, W., Zapf, J. and Froesch, E.R. (1986) *Nature* **323**, 169-170

Scholtz, J.M. and Robertson, A.D. (1995) Hydrogen Exchange Techniques, In: *Methods in Molecular Biology* **40** ed. Shirley, B.A. pp 291-311

Schurr, J.M., babcock, H.P. and Fujimoto, B.S. (1994) *J. Magn. Reson.* **B105**, 211-224

Schäffer, L., Larsen, U.D., Linde, S., Hejnaes, K.R. and Skriver, L. (1993) *Biochem. Biophys. Acta* **1203**, 205-209

Shimizu, M. (1986) *Am. J. Physiol.* **251**, E611

Shooter, G.K., Magee, B., Soos, M.A., Francis, G.L., Siddle, K. and Wallace, J.C. (1996) *J. Mol. Endocrinol.* **17**, 237-246

Shortle, D. (1989) *J. Biol. Chem.* **264**, 5315-5318

Sklar, M.M., Kiess, W., Thomas, C.L. and Nissley, S.P. (1989) *J. Biol. Chem.* **264**, 16733

Sklenár, V. Piotto, M., Leppik, R. and Saudek, V. (1993) *J. Magn. Reson.* **102**, 241-245

Sommer, A., Spratt, S.K., Tatsuno, G.P., Tressel, T., Lee, R. and Maack, C.A. (1993) *Growth Regul.* **3**, 46-49

Soos, M.A., Whittaker, J., Lammers, R., Ullrich, A. and Siddle, K. (1990) *Biochem. J.* **270**, 383-390

Spencer, E.M. and Chan, K. (1995) *Prog. Growth Factor Res.* **6**, 209-214

Spencer, E.M., Skover, G., and Hunt, T.K. (1988) In: *Growth factors and other aspects of wound healing: Biological implications*, Liss Inc., New York, ed. R. Alan

Spera, S., Ikura, M. and Bax, A. (1991) *J. Biomol. NMR* **1**, 155-165

Steele, N.C. and Elsasser, T.H. (1989) In *Animal Growth Regulation*, Plenum Press, New York, pp 295-316

Stone, M.J., Fairbrother, W.J., Palmer, A.G., III, Reizer, J., Saier, M.H., Jnr, and Wright, P.E. (1992) *Biochemistry* **31**, 4394-4406

Stone, M.J., Chandrasekhar, K., Holmgren, A., Wright, P.E. and Dyson, H.J. (1993) *Biochemistry* **32**, 426-435

Sturtevant, J.M. (1994) *Curr. Opin. Struct. Biol.* **4**, 69-78

Svoboda, M.E., Van Wyk, J.J, Klapper, D.G., Fellows, R.E., Grissom, F.E. and Schlueter, R.J. (1980) *Biochemistry* **19**, 790-797

Swisshelm, K., Ryan, K., Tsuchiya, K. and Sager, R. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 4472-4476

Szabo, L., Mottershead, D.G., Ballard, F.J. and Wallace, J.C. (1988) *Biochem. Biophys. Res. Commun.* **151**, 207-214

Szyperski, T., Güntert, P., Otting, G. and Wüthrich, K. (1992) *J. Magn. Reson.* **99**, 552-560

Tally, M., Li, C.H. and Hall, K. (1987) *Biochem. Biophys. Res. Commun.* **148**, 811-816

Tartare, S., Mothe, I., Kowalski-Chauvel, A., Breittmayer, J-P., Ballotti, R. and Van Obberghen, E. (1994) *J. Biol. Chem*, **269**, 11449-11455

Terasawa, H., Kohda, D., Hatanaka, H., Nagat, K., Migashihashi, H., Sakano, K. and Inagaki, F. (1994) *EMBO J.* **13**, 5590-5597

Tjandra, N., Feller, S.E., Pastor, R.W. and Bax, A. (1995) *J. Am. Chem. Soc.* **117**, 12562-12566

Torres, A.M., Forbes, B.E., Aplin, S.E., Wallace, J.C. Francis, G.L. and Norton, R.S. (1995) *J. Mol. Biol.* **248**, 385-401

Twigg, S.M. and Baxter, R.C. (1998) *J. Biol. Chem.* **273**, 6074-6079

Upton, F.Z., Szabo, L., Wallace, J.C. and Ballard, F.J. (1990) *J. Mol. Endocrinol.* **5**, 77-84

Voutilainen, R. and Miller, W.L. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1590-1594

Vuister, G.W. and Bax, A. (1993) *J. Am. Chem. Soc.* **115**, 7772-7777

Vuister, G.W., Boelens, R., Kaptein, R., Burgering, M. and Van Zilj, P.C.M. (1992) *J. Biomol. NMR* **2**, 301-305

Wagner, G. (1997) An account of NMR in structural biology, NMR Supplement, In: *Nature Struct. Biol.* **4**, 841-865

Walsh, G. (1995) *Biotechnology* **13**, 1167-1171

Walton, P., Wallace, J.C. and Ballard, J.F. (1990) *Today's Life Science*, April, 12-19

Weissman, J.S. and Kim, P.S. (1992) *Cell*, **71**, 841-851

Werner, H. and De Roth, D. (1988) *Adv. Cancer Res.* **68**, 183-223

Williamson, M.P., Havel, T.F. and Wüthrich, K. (1985) *J. Mol. Biol.* **182**, 295-315

Winter, J.R. and Sorenson, P. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9330-9334

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995a) *J. Biomol. NMR* **5**, 67-81

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.K. and Sykes, B.D. (1995b) *J. Biomol. NMR* **6**, 135-140

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1992) *Biochemistry* **31**, 1647-1651

Wishart, D.S. and Sykes, B.D. (1994) *Methods Enzymol.* **239**, 363-392

Woessner, D.T. (1962) *J. Chem. Phys.* **37**, 647-654

Wüthrich, K. (1989) *Acc. Chem. Res.* **22**, 36-44

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York

Wüthrich, K., Wider, G., Wagner, G. and Braun, W. (1982) *J. Mol. Biol.* **155**, 311-319

Yeh, L.C., Adamo, M.L., Olson, M.S. and Lee, J.C. (1997) *Endocrinology* **138**, 4181-4190

Zapf, J. (1995) *Eur. J. Endocrinol.* **132**, 645-654

Zapf, J., Kiefer, M., Merryweather, J., Musiarz, F., Bauer, D., Born, W., Fischer, J.A. and Froesch, E.R. (1990) *J. Biol. Chem.* **265**, 14892-14898

Zapf, J., Schoenle, E., Jagars, G., Sand, I., Grunwald, J. and Froesch, E.R. (1979) *J. Clin. Invest.* **63**, 1077-1084

Zhang, O., Kay, L.E., Olivier, J.P. and Forman-Kay, J.D. (1994) *J. Biomol. NMR* **4**, 845-858

Zheng, Z., Czaplicji, J. and Jardetzky, O. (1995) *Biochemistry* **34**, 5212

Zhou, N.E., Zhu, B-Y., Sykes, B.D. and Hodges, R.S. (1992) *J. Am. Chem. Soc.* **114**, 4320-4326

Zhu, X.L., Ohta, Y., Jordan, F. and Inouye, M. (1989) *Nature*, **339**, 483-484

APPENDIX 1

NMR Pulse Sequences

Pulse sequences are shown for each of the NMR experiments performed with the exception of the following; 2D NOESY and 2D TOCSY which are standard experiments, 2D ^{15}N-HSQC-TOCSY and 2D ^{15}N-HSQC-NOESY which are 2D variants of the 3D experiments shown, and the 3D ^{15}N-HSQC-NOESY-HSQC. The same pulse scheme was used for the 3D ^{15}N-HSQC-NOESY-HSQC and the 3D ^{15}N-HSQC-TOCSY-NOESY-HSQC with the TOCSY element disabled while acquiring the 3D ^{15}N-HSQC-NOESY-HSQC spectrum.

Pulses for which the phases are not labeled are applied along the x-axis. Pulses labeled with $\pi/2$ represent a flip angle of 90°, while pulses labeled with π correspond to a flip angle of 180°. Water saturation/dephasing was minimised through the use of a one-lobe H_2O Sinc pulse (shaped pulse in Figures). The ^{1}H carrier was placed on the water resonance for the duration of the experiment. The DIPSI mixing scheme was used for all TOCSY experiments. The ^{15}N pulses were applied with a field of kHz, with nitrogen decoupling during acquisition achieved through the use of a kHz WALTZ decoupling field. The delays τ, $\delta1$ and $\delta2$ were set to $1/(4J_{NH}) = 2.3$, 1.5 and 0.5 ms respectively.

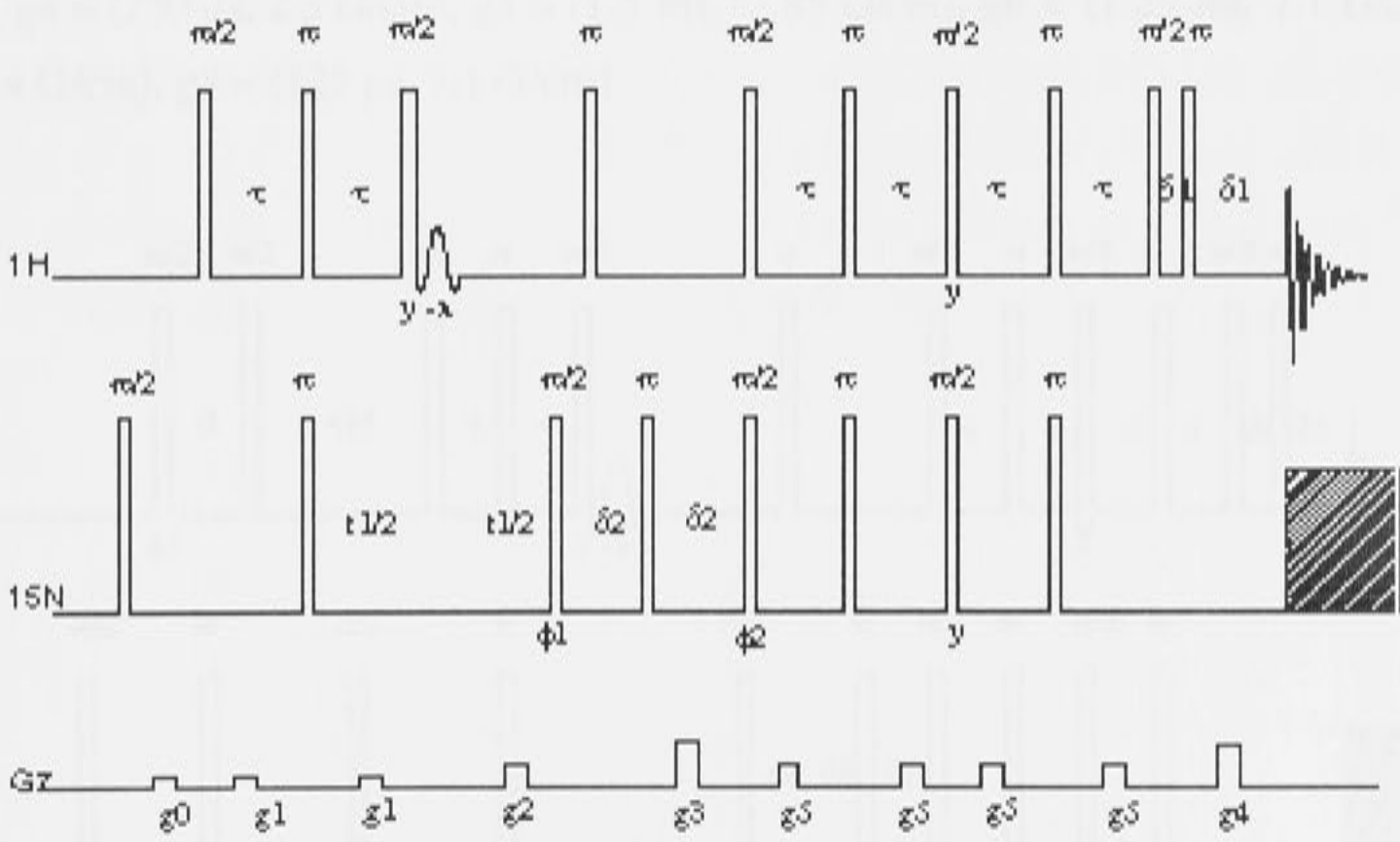

Figure 1: 2D ^{15}N-HSQC, ϕ_1 = (x,-x); ϕ_2 = x. The gradient strengths and durations are: g0 = (1 ms, 4 G/cm), g1 = (500 µs, 4 G/cm), g2 = (500 µs, 7.5 G/cm), g3 = (1.25 ms, 15 G/cm), g4 = (125 µs, 14.1 G/cm), g5 = (150 µs, 7.4 G/cm).

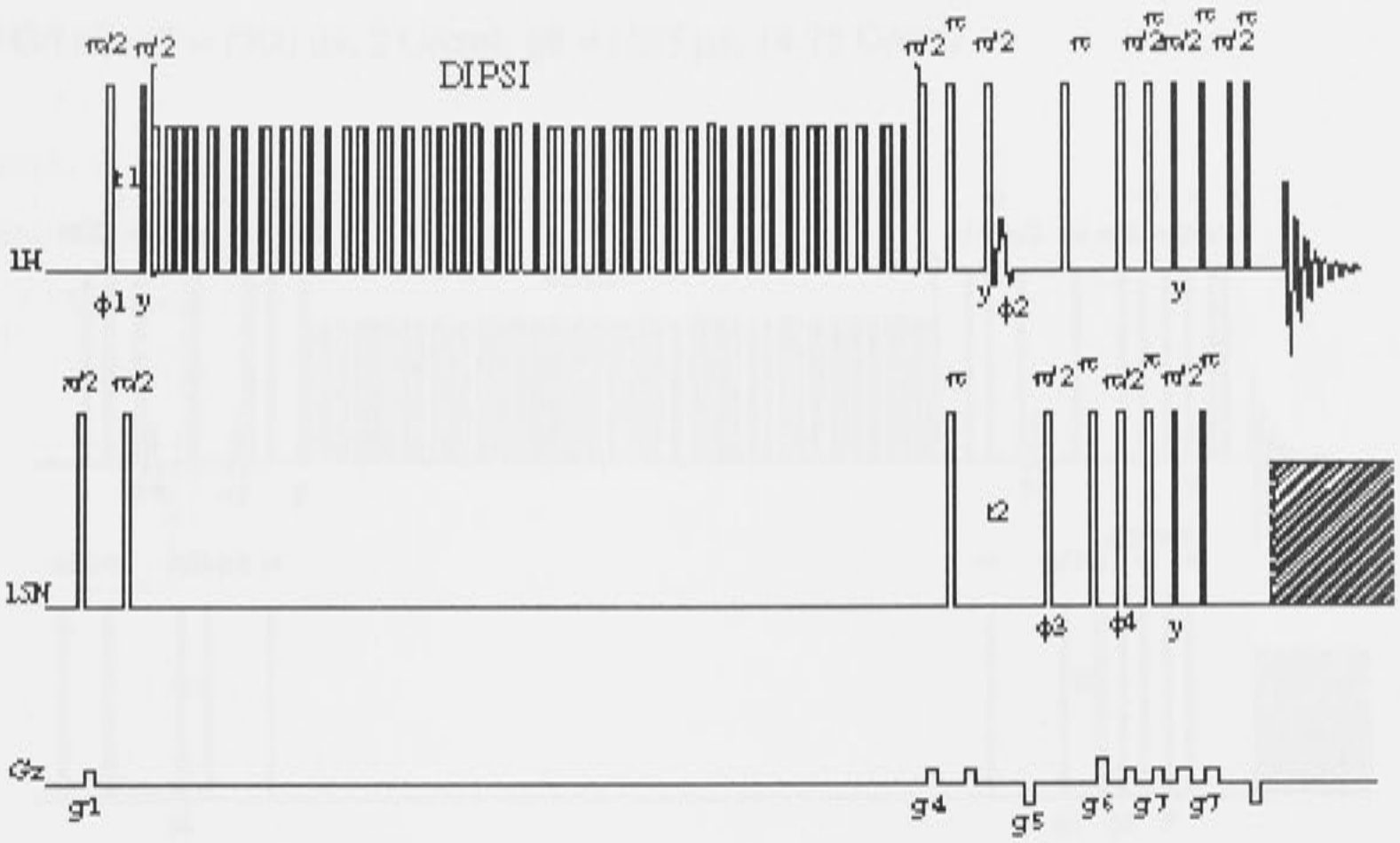

Figure 2: 3D ^{15}N-TOCSY-HSQC, ϕ1 = (y,-y); ϕ2 = (x,-x); ϕ3 = 2(x), 2(-x); ϕ4 = x. The gradient strengths and durations are: g1 = (1 ms, 2 G/cm), g2 = (1 ms, 2.5 G/cm), g3 = (500 µs, 2 G/cm), g4 = (750 µs, 2.5 G/cm), g5 = (1.5 ms, -7.85 G/cm), g6 = (1.25 ms, 7.1 G/cm), g7 = (500 µs, 4 G/cm), g8 = (125 µs, 7.1 G/cm)

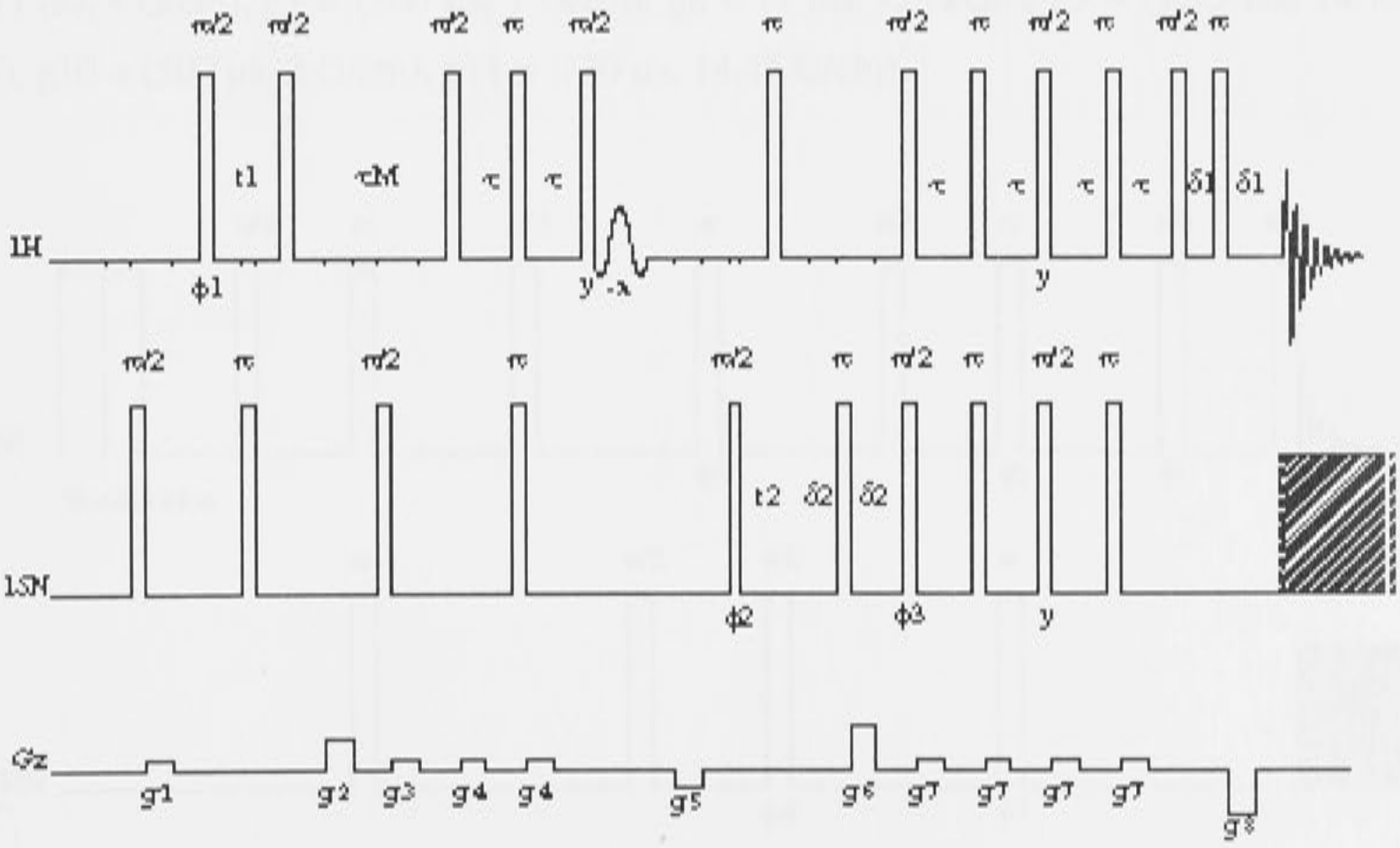

Figure 3: 3D ^{15}N-NOESY-HSQC, ϕ1 = (135°, 315°); ϕ2 = 2(x), 2(-x); ϕ3 = x; t_M = mixing time. The gradient strengths and durations are: g1 = (1ms, 2 G/cm), g2 = (1 ms, 10 G/cm), g3 = (1ms, 4 G/cm), g4 = (500 µs, 4 G/cm), g5 = (1ms, -5 G/cm), g6 = (1.25 ms,

14.75 G/cm), g7 = (500 µs, 2 G/cm), g8 = (125 µs, 14.75 G/cm)

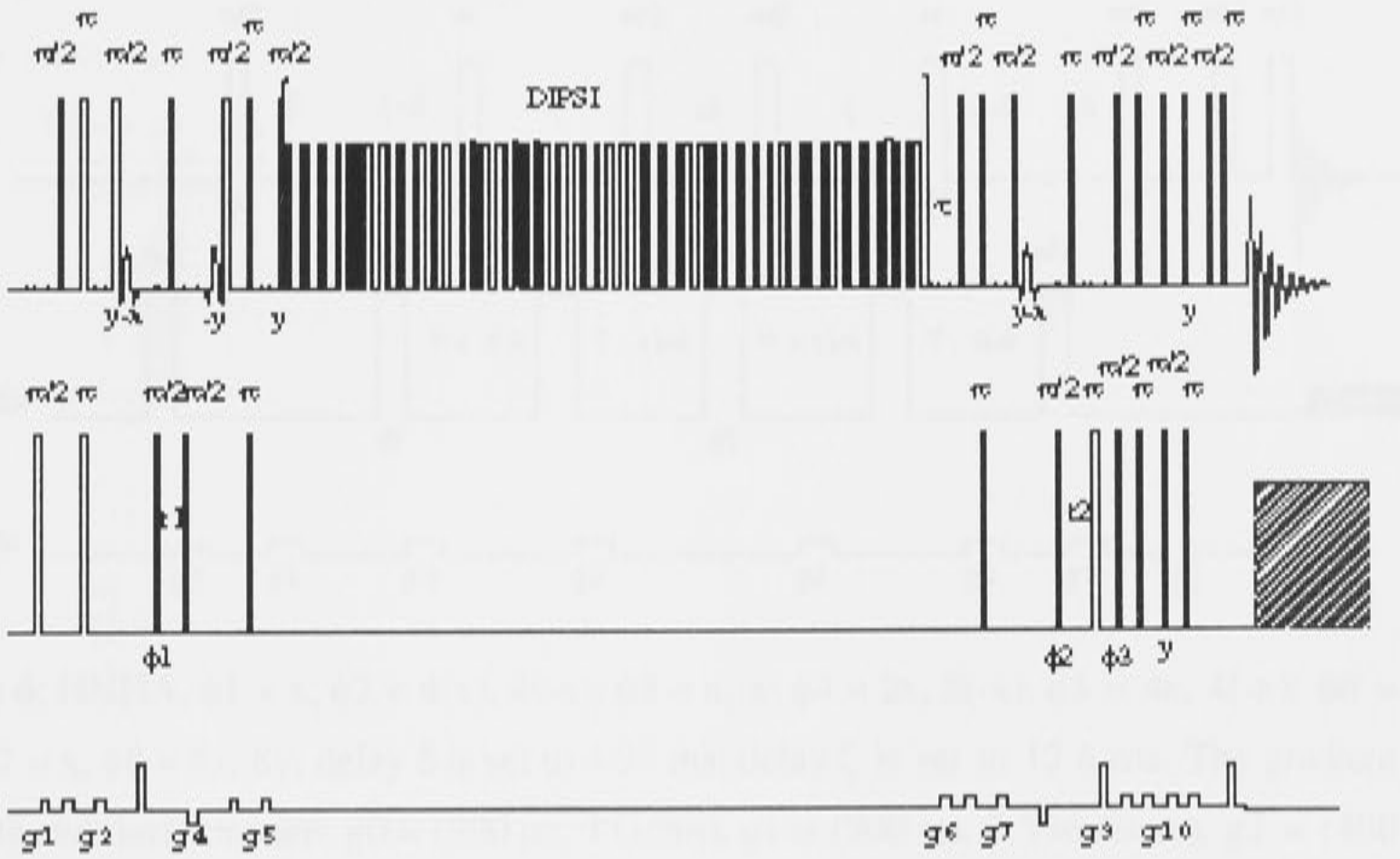

Figure 4: 3D ^{15}N-HSQC-TOCSY-NOESY-HSQC, ϕ1 = (x,-x); ϕ2 = 2(x), 2(-x); ϕ3 = x, t = mixing time. The gradient strengths and durations are: g1 = (1 ms, 2 G/cm), g2 = (500 µs, 1 G/cm), g3 = (1 ms, 7.5 G/cm), g4 = (500 µs, -3.5 G/cm), g5 = (500 µs, 2.5 G/cm), g6 = (1 ms, 4 G/cm), g7 = (500 µs, 1 G/cm), g8 = (1 ms, -5 G/cm), g9 = (1.25 ms, 14.7 G/cm), g10 = (500 µs, 2 G/cm), g11 = (130 µs, 14.45 G/cm)

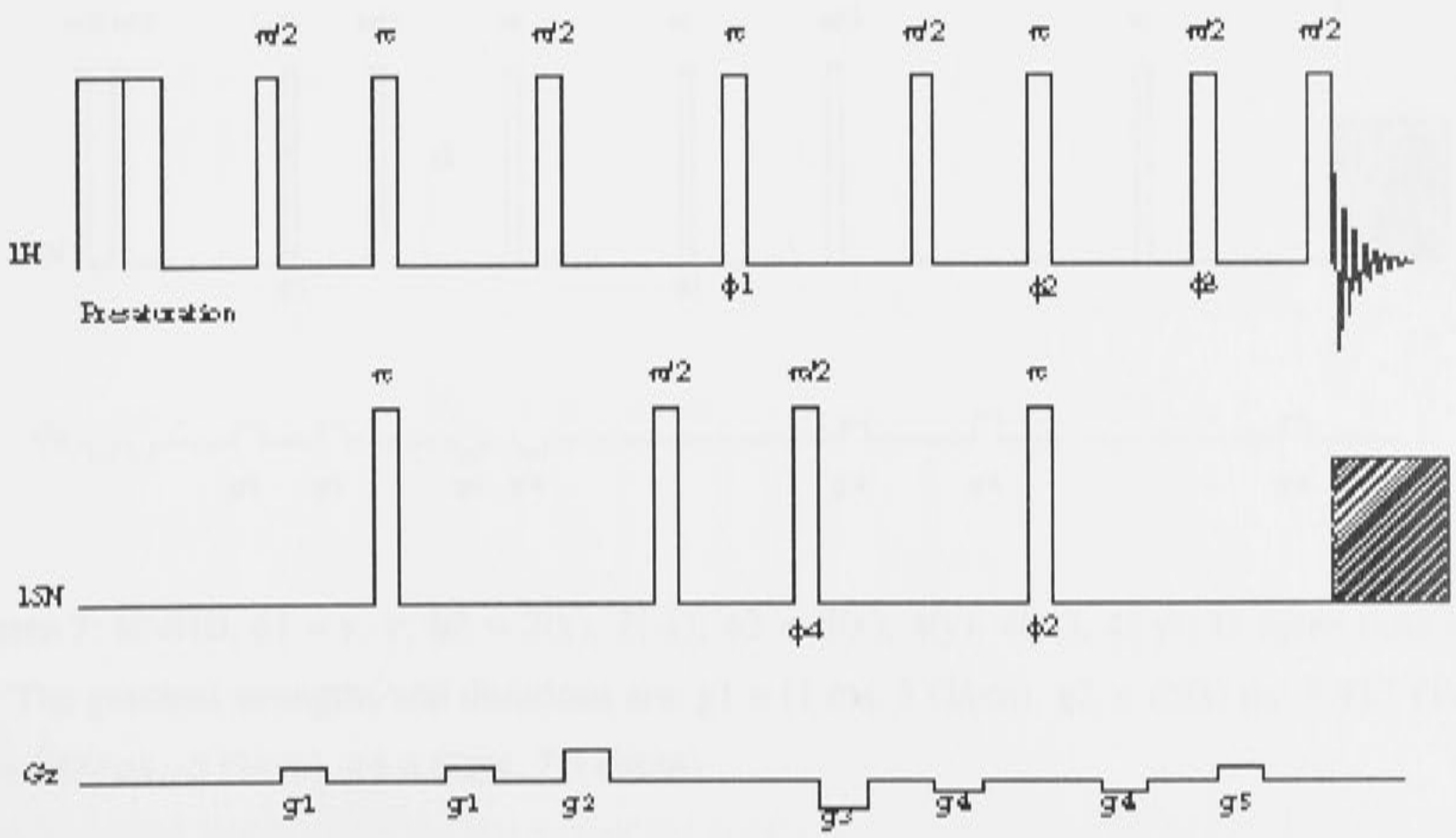

Figure 5: HMQC-J, $\phi1 = 8(x,-x)$; $\phi2 = 4(x,x,-x,-x)$; $\phi3 = 2(x,x,x,x,-x,-x,-x,-x)$; $\phi4 = 8(x), 8(-x)$.

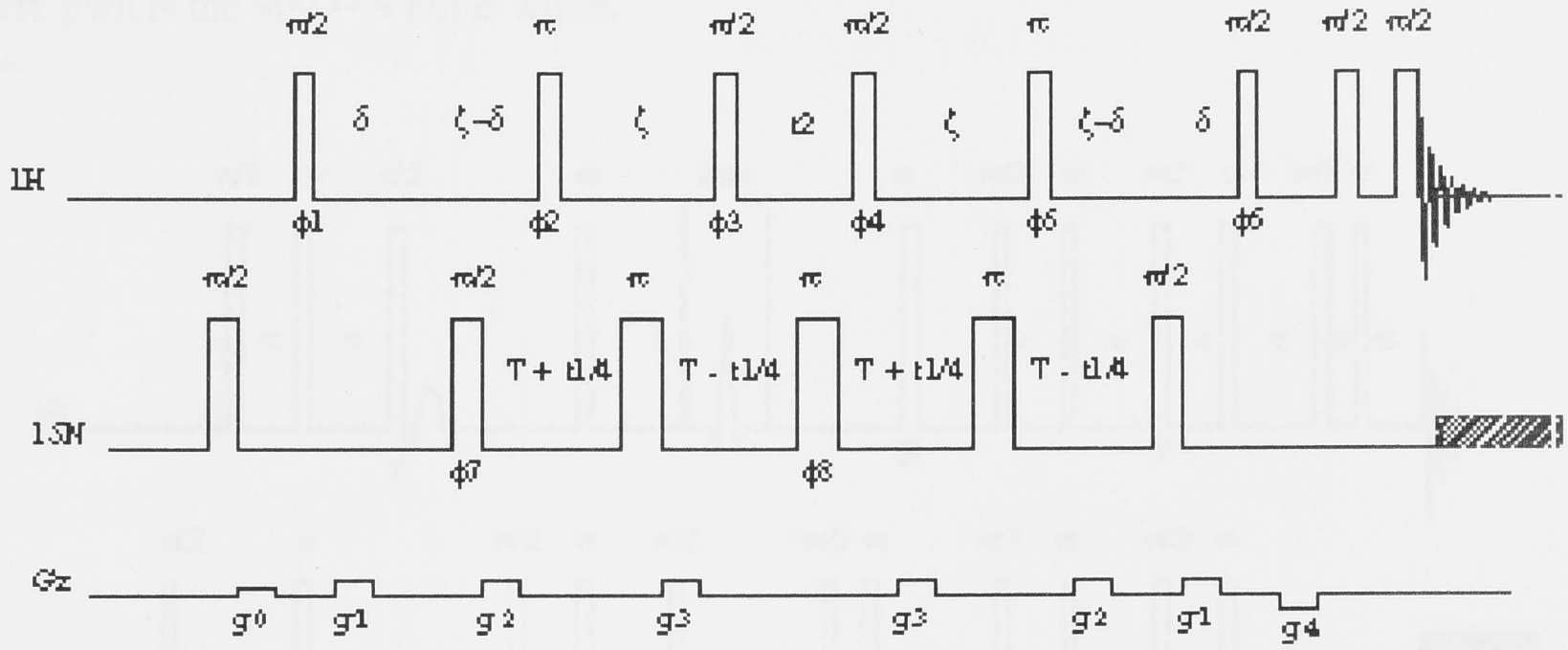

Figure 6: HNHA, $\phi1 = x$; $\phi2 = 4(x), 4(-x)$; $\phi3 = x,-x$; $\phi4 = 2x, 2(-x)$; $\phi5 = 4x, 4(-x)$; $\phi6 = y,-y$; $\phi7 = x$, $\phi8 = 8x, 8y$; delay δ is set to 4.95 ms; delay ζ is set to 12.6 ms. The gradient stengths and durations are: g0 = (500 µs, 4 G/cm), g1 = (300 µs, 7.396 G/cm), g2 = (400 µs, 7.396 G/cm), g3 = (750 µs, 7.4 G/cm), g4 = (500 µs, -7.4 G/cm)

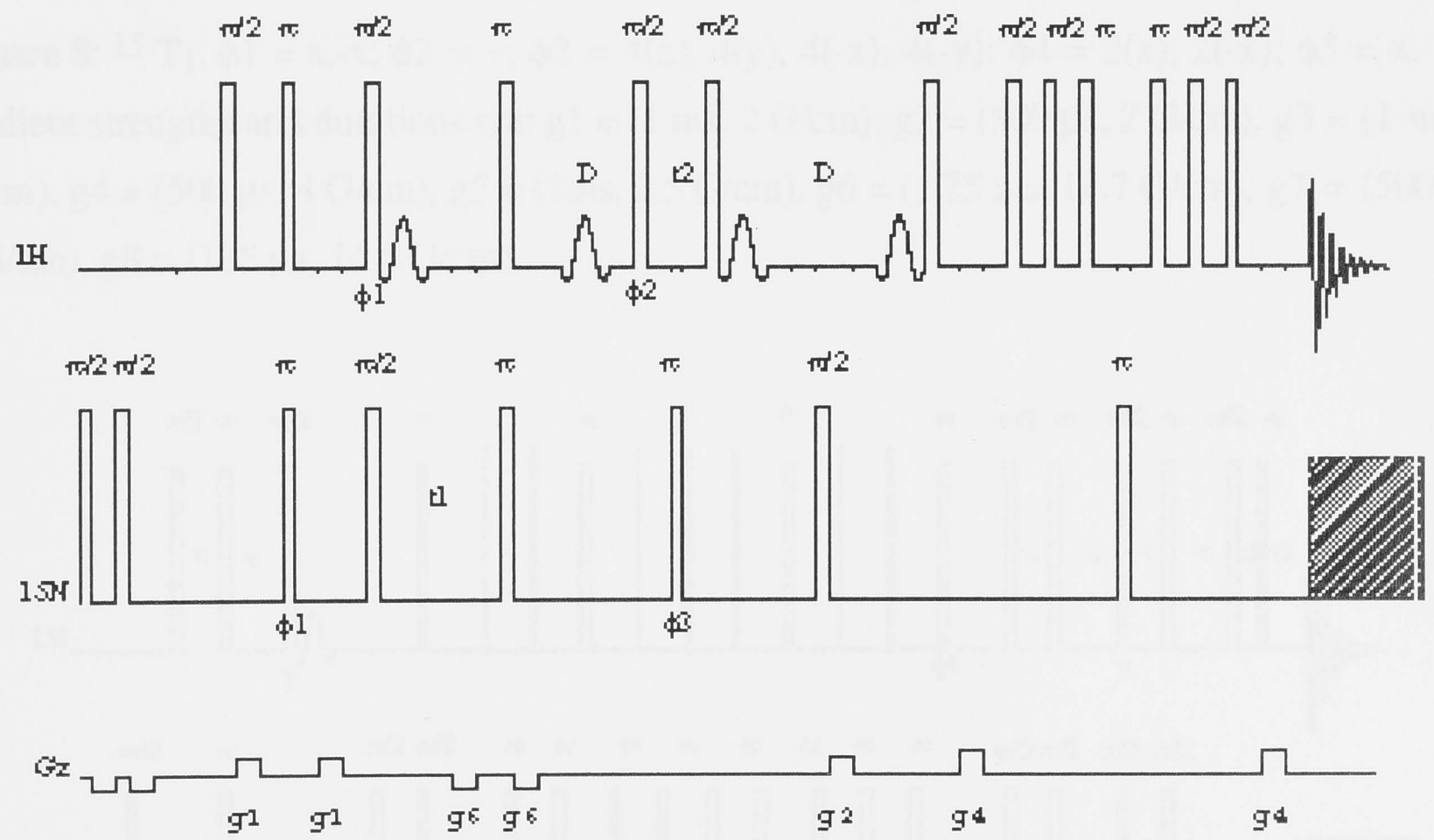

Figure 7: HNHB, $\phi1 = y,-y$; $\phi2 = 2(x), 2(-x)$; $\phi3 = 4(x), 4(y), 4(-x), 4(-y)$; D delay time = 38 ms. The gradient strengths and durations are: g1 = (1 ms, 5 G/cm), g2 = (500 µs, 5.917 G/cm), g3 = (500 µs, -5 G/cm), g4 = (1ms, 7.5 G/cm)

For the following three experiments; ^{15}N T_1, ^{15}N T_2 and ^{15}N NOEs, τ = 2.25 ms ($1/(4J_{NH})$);

τ_A = 2.75 ms (1/(4J_{NH})); τ_B = 0.5 ms. The delay τ'_A is set according to the relation : $\tau'_A = \tau_A$ + 2pw where pw is the width of the ^{1}H 90^o pulse. The delay τ''_A is set as $\tau''_A = \tau_A$ + (2/p) pwn where pwn is the 90^o ^{15}N pulse width.

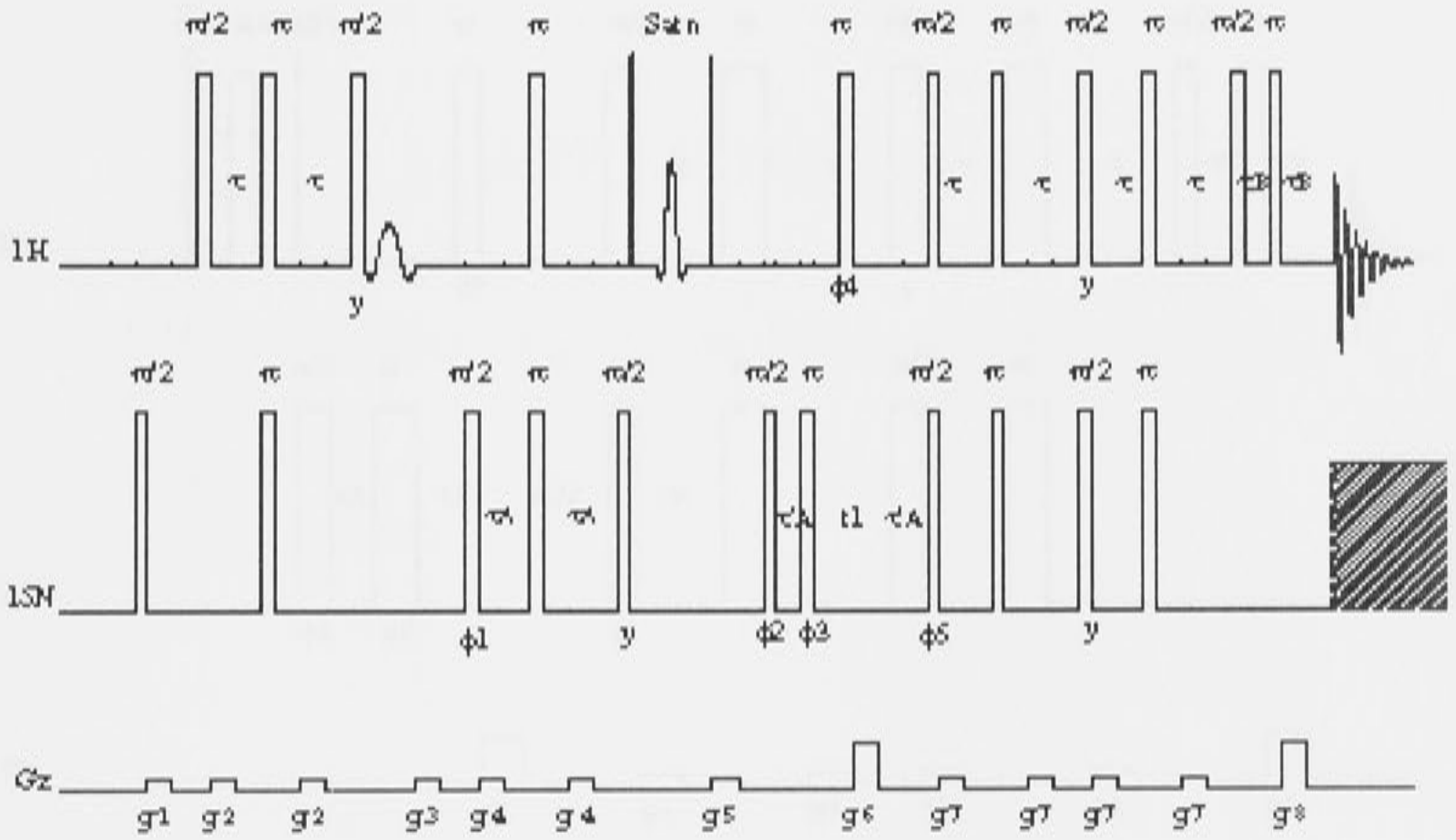

Figure 8: ^{15} T_1, ϕ1 = x,-x; ϕ2 = y; ϕ3 = 4(x), 4(y), 4(-x), 4(-y); ϕ4 = 2(x), 2(-x); ϕ5 = x. The gradient strengths and durations are: g1 = (1 ms, 2 G/cm), g2 = (500 µs, 2 G/cm), g3 = (1 ms, 4 G/cm), g4 = (500 µs, 4 G/cm), g5 = (1ms, 2.5 G/cm), g6 = (1.25 ms, 14.7 G/cm), g7 = (500 µs, 2 G/cm), g8 = (125 µs, 14.6 G/cm)

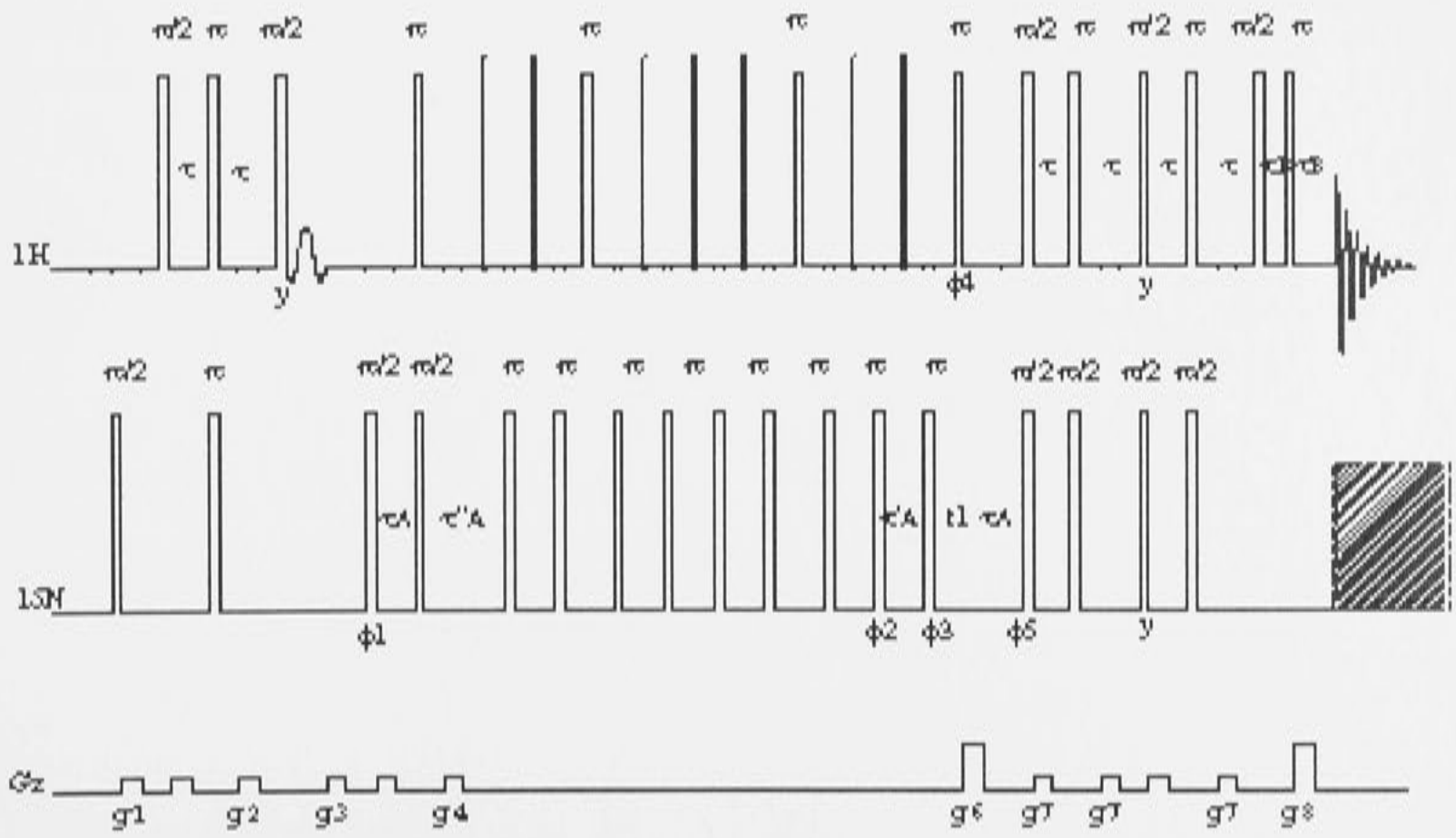

Figure 9: ^{15}N T_2, $\phi 1$ = x,-x; $\phi 2$ = y; $\phi 3$ = 4(x), 4(y), 4(x). The gradient strength and durations are: g1 = (1 ms, 2 G/cm), g2 = (500 μs, 2 G/cm), g3 = (1 ms, 4 G/cm), g4 = (500 μs, 4 G/cm), g6 = (1.25 ms, 14.7 G/cm), g7 = (500 μs, 2 G/cm), g8 = (125 μs, 14.6 G/cm)

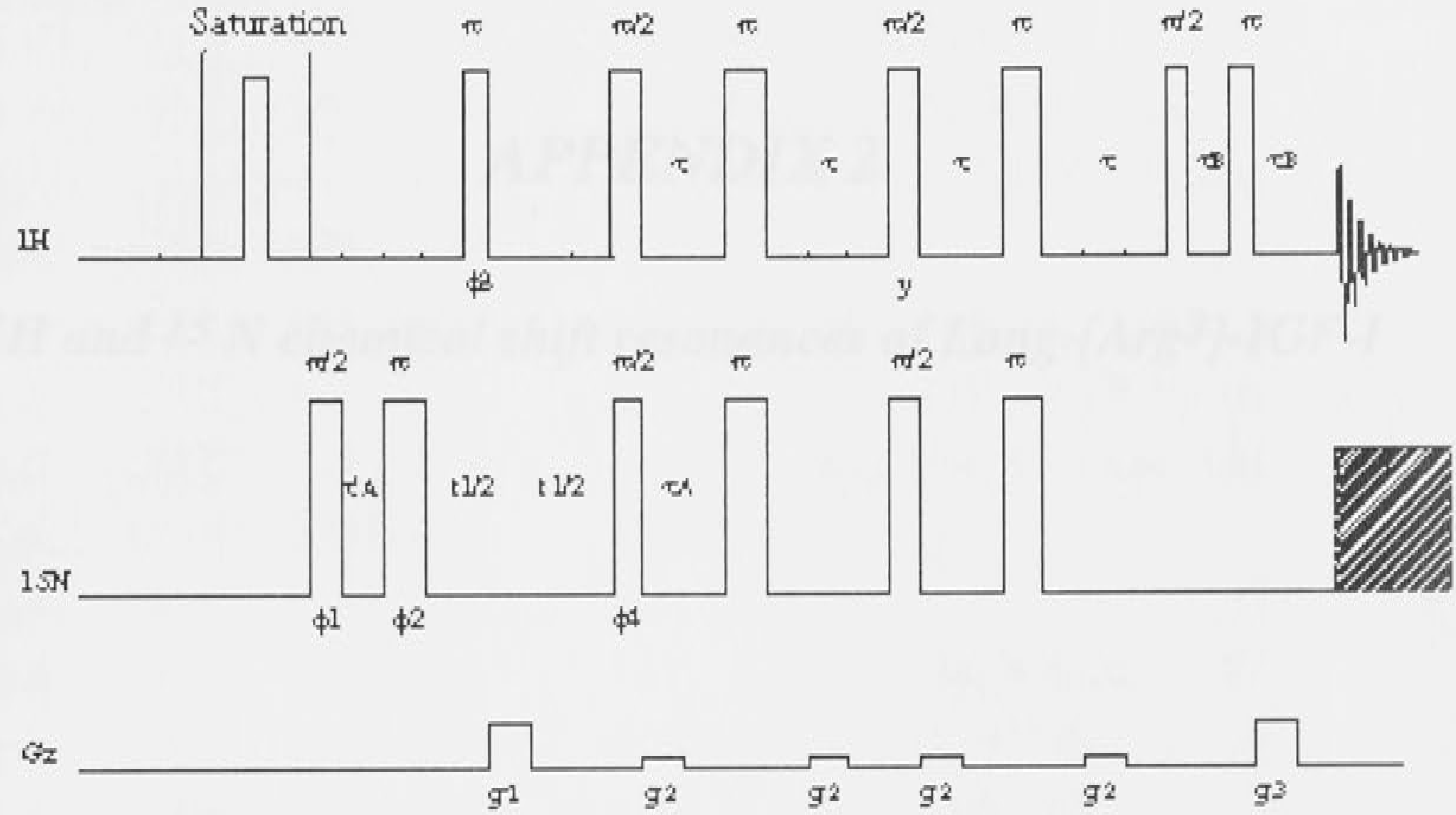

Figure 10: ^{15}N NOE, $\phi 1$ = y; $\phi 2$ = 2(x), 2(y), 2(-x), 2(-y); $\phi 3$ = x,-x; $\phi 4$ = x. The gradient strengths and durations are: g1 = (1.25 ms, 14.65 G/cm), g2 = (500 μs, 2 G/cm), g3 = (125 μs, 14.65 G/cm)

APPENDIX 2

^{1}H and ^{15}N chemical shift resonances of Long-[Arg^{3}]-IGF-I

Assignments for the ^{1}H and ^{15}N resonances of Long-[Arg^{3}]-IGF-1 in H_2O at pH 3.0 and 30 °C

n.o. = not observed

Amino acid	Chemical shift of				
	^{15}N	NH	H^{α}	H^{β}	Others
Met -13	121.9	8.22	4.35	2.07, 2.05	$^{\gamma}CH_2$ 2.64, 2.57; $^{\varepsilon}CH_3$ 1.95
Phe -12	125.8	8.65	4.75	3.20, 2.98	$^{\delta}H$ 7.26; $^{\varepsilon}H$ n.o.; $^{\zeta}H$ n.o.
Pro -11			4.40	2.29	$^{\gamma}CH_2$ 2.07,2.09; $^{\delta}CH_2$ n.o.
Ala -10	126.3	8.36	4.41	1.41	
Met -9	121.9	8.24	4.34	2.06, 2.02	$^{\gamma}CH_2$ 2.61, 2.52; $^{\varepsilon}CH_3$ 1.90
Pro -8			4.45	2.17	$^{\gamma}CH_2$ 1.89, 1.94; $^{\delta}CH_2$ 3.63, 3.75
Leu -7	122.1	8.22	4.33	1.80	$^{\gamma}CH$ 1.84; $^{\delta}CH_3$ 1.41, 1.41
Ser -6	119.3	8.23	4.50	3.99, 3.91	
Ser -5	119.5	8.31	4.43	3.95, 3.90	
Leu -4	129.2	7.92	4.29	1.61	$^{\gamma}CH$ 1.38; $^{\delta}CH_3$ 0.92,0.87
Phe -3	119.7	8.35	5.04	3.06, 2.69	$^{\delta}H$ 7.22; $^{\varepsilon}H$ 7.00; $^{\zeta}H$ n.o.
Val -2	121.6	7.40	4.05	2.19	$^{\gamma}CH_3$ 1.04, 1.00
Asn -1	124.3	8.49	4.15	2.86, 2.74	$^{\gamma}NH_2$ n.o.
Gly 1	110.9	8.13	4.05, 3.64		
Pro 2			4.39	2.31	$^{\gamma}CH_2$ 2.05; $^{\delta}CH_2$ 3.76, 3.79
Arg 3	119.3	8.56	4.51	2.07, 1.80	$^{\gamma}CH_2$ 1.92, 1.51; $^{\delta}CH_2$ 3.62, 3.66; $^{\varepsilon}NH$ 7.21
Thr 4	115.3	8.07	4.42	4.30	$^{\gamma}CH_3$ 1.25
Leu 5	123.7	8.31	4.23	1.58	$^{\gamma}CH$ 1.61; $^{\delta}CH_3$ 0.91, 0.95
Cys 6	119.7	8.38	4.58	3.15, 2.88	
Gly 7	110.0	7.74	3.99, 3.70		
Ala 8	130.3	8.98	3.92	1.44	
Glu 9	118.1	8.00	4.04	2.26, 2.07	$^{\gamma}CH_2$ 2.78, 2.52
Leu 10	121.7	6.88	3.90	1.89	$^{\gamma}CH$ 2.07; $^{\delta}CH_3$ 1.05, 0.76
Val 11	119.3	7.29	3.28	2.10	$^{\gamma}CH_3$ 0.99, 0.99
Asp 12	128.1	8.09	4.49	2.82, 2.71	
Ala 13	124.1	7.92	4.07	1.44	
Leu 14	120.5	8.11	3.76	1.14	$^{\gamma}CH$ 1.32; $^{\delta}CH_3$ 0.64, 0.39
Gln 15	119.9	8.34	4.14	2.29, 2.10	$^{\gamma}CH_2$ 2.62, 2.47, $^{\delta}NH_2$ 7.29, 6.77
Phe 16	120.6	7.84	4.08	3.29, 3,24	$^{\delta}H$ 7.13; $^{\varepsilon}H$ 7.25; $^{\zeta}H$ n.o.

Val 17	119.2	8.70	3.61	2.10	$^{\gamma}CH_3$ 0.99, 1.15
Cys 18	117.1	8.67	4.81	3.34, 2.91	
Gly 19	110.4	7.72	3.99		
Asp 20	125.3	8.85	4.51	2.87, 2.82	
Arg 21	119.7	8.11	4.17	2.07, 2.02	$^{\gamma}CH_2$ 1.89, 1.79; $^{\delta}CH_2$ 3.33, 3.33; $^{\varepsilon}NH$ 7.14
Gly 22	105.0	7.38	4.04, 3.76		
Phe 23	115.2	7.53	5.18	3.24, 2.91	$^{\delta}H$ 6.79; $^{\varepsilon}H$ 6.81; $^{\zeta}H$ n.o.
Tyr 24	120.7	8.66	4.82	3.21, 2.93	$^{\delta}H$ 7.13, $^{\varepsilon}H$ 6.81
Phe 25	119.6	8.48	4.79	3.21, 3.02	$^{\delta}H$ 7.22; $^{\varepsilon}H$ 7.00; $^{\zeta}H$ n.o.
Asn 26	119.8	8.18	5.00	2.86, 2.71	$^{\gamma}NH_2$ 7.54, 6.90
Lys 27	123.9	8.43	4.44	1.74, 1.70	$^{\gamma}CH_2$ 1.37, 1.37; $^{\delta}CH_2$ 1.65, 1.65; $^{\varepsilon}CH_2$ 2.83, 2.73
Pro 28			4.42	2.25	$^{\gamma}CH_2$ 1.92, 1.93; $^{\delta}CH_2$ 3.59, 3.71
Thr 29	115.3	8.15	4.21	4.32	$^{\gamma}CH_3$ 1.21
Gly 30	111.8	8.30	3.95, 3.90		
Tyr 31	121.6	8.11	4.54	3.06, 2.93	$^{\delta}H$ 7.39; $^{\varepsilon}H$ 6.75
Gly 32	112.5	8.39	3.99, 3.87		
Ser 33	117.2	8.20	4.46	3.94, 3.85	
Ser 34	118.6	8.19	4.43	3.90, 3.84	
Ser 35	118.8	8.38	4.48	3.97, 3.98	
Arg 36	123.7	8.16	4.33	1.87, 1.85	$^{\gamma}CH_2$ 1.75, 1.54; $^{\delta}CH_2$ 3.17, 3.17; $^{\varepsilon}NH$ n.o.
Arg 37	122.6	8.15	4.31	1.82, 1.81	_CH2 1.69, 1.58; _CH2 3.14, 3.14; _NH n.o.
Ala 38	127.5	8.12	4.57	1.35	
Pro 39			4.39	2.33	_CH2 2.02, 2.04; _CH2 3.70, 3.74
Gln 40	121.5	8.48	4.40	2.03, 1.99	_CH2 2.40, 2.26; _NH2 7.48, 6.81
Thr 41	115.6	8.09	4.44	4.29	_CH3 1.25
Gly 42	113.2	8.60	4.30, 4.16		
Ile 43	122.4	7.89	3.81	1.11	_CH2 0.95, 0.90; _CH3 0.72; _CH3 0.63
Val 44	123.59	8.04	3.55	1.93	_CH3 0.91, 0.71
Asp 45	121.8	7.81	4.44	2.82, 2.78	
Glu 46	119.9	8.24	4.17	2.06, 2.01	_CH2 1.88 1.80
Cys 47	113.6	8.26	5.04	3.27, 3.03	
Cys 48	117.2	7.36	4.44	3.18, 2.91	
Phe 49	117.3	7.72	4.69	3.59, 3.43	_H 7.39; _H 7.23; _H n.o.

Arg 50	119.9	7.53	4.48	1.93	_CH2 1.71, 1.71; _CH2 3.22, 3.22; _NH 7.34
Ser 51	113.3	7.83	4.24	3.72, 3.62	
Cys 52	117.7	8.01	4.70	3.56, 3.25	
Asp 53	118.3	7.95	4.23	2.93, 2.78	
Leu 54	123.6	7.59	3.92	1.36	$^{\gamma}$CH 1.32; $^{\delta}CH_3$ 0.90, 0.70
Arg 55	119.9	7.86	4.16	2.05, 1.85	$^{\gamma}CH_2$ 1.63, 1.54; $^{\delta}CH_2$ 3.37, 3.22; $^{\varepsilon}$NH n.o.
Arg 56	117.7	7.80	3.93	1.83, 1.67	$^{\gamma}CH_2$ 1.61, 1.56; $^{\delta}CH_2$ 3.16, 2.82; $^{\varepsilon}$NH 7.32
Leu 57	124.3	8.30	4.42	1.46	$^{\gamma}$CH 1.64; $^{\delta}CH_3$ 0.85, 0.80
Glu 58	114.8	8.09	4.20	1.85, 1.53	$^{\gamma}CH_2$ 2.17, 2.07
Met 59	118.9	7.61	4.21	2.16, 1.99	$^{\gamma}CH_2$ 2.84, 2.58; $^{\varepsilon}$CH 1.64
Tyr 60	117.0	7.98	4.72	2.82, 2.78	$^{\delta}$H 7.37; $^{\varepsilon}$H 6.74
Cys 61	116.9	7.26	5.15	3.21, 2.94	
Ala 62	128.6	8.35	4.36	1.22	
Pro 63			4.42	2.27	$^{\gamma}CH_2$ 2.07, 2.07; $^{\delta}CH_2$ 3.70, 3.84
Leu 64	123.9	8.30	4.42	1.56	$^{\gamma}$CH 1.87; $^{\delta}CH_3$ 0.93, 0.83
Lys 65	125.5	8.34	4.61	1.75, 1.60	$^{\gamma}CH_2$ 1.46, 1.56; $^{\delta}CH_2$ 1.70, 1.70
Pro 66			4.45	2.16	$^{\gamma}CH_2$ 2.02, 2.07; $^{\delta}CH_2$ 3.62, 3.82
Ala 67	125.7	8.18	4.29	1.39	
Lys 68	123.3	8.02	4.17	1.89, 1.85	$^{\gamma}CH_2$ 1.41, 1.37; $^{\delta}CH_2$ n.o.; $^{\varepsilon}CH_2$ n.o.
Ser 69	113.8	8.01	4.23	4.04, 3.90	
Ala 70	132.1	8.03	4.19	1.36	

APPENDIX 3

^{1}H and ^{15}N chemical shift resonances of Long-[Leu^{60}]-IGF-I

Assignments for the ^{1}H and ^{15}N resonances of Long-[Leu^{60}]-IGF-1 in H_2O at pH 3.0 and 30 °C

n.o. = not observed

Amino acid	Chemical shift of				
	^{15}N	NH	H^{α}	H^{β}	Others
Met -13	120.3	8.17	4.25	1.76, 1.82	$^{\gamma}CH_2$ 1.40; $^{\varepsilon}CH_3$ n.o.
Phe -12	124.2	8.55	4.87	3.10, 2.89	$^{\delta}H$ 7.10; $^{\varepsilon}H$ n.o.; $^{\zeta}H$ n.o.
Pro -11			n.o.	2.00, 1.92	$^{\gamma}CH_2$ 1.84, 1.79; $^{\delta}CH_2$ 3.62
Ala -10	124.4	8.30	4.23	1.33	
Met -9	120.3	8.04	4.65	1.84	$^{\gamma}CH_2$ 2.45; $^{\varepsilon}CH_3$ 1.92
Pro -8			4.41	2.20	$^{\gamma}CH_2$ 1.92, 1.83; $^{\delta}CH_2$ 3.24
Leu -7	123.6	8.11	4.22	1.31	$^{\gamma}CH$ 1.16; $^{\delta}CH_3$ 0.99, 0.76
Ser -6	117.3	8.15	4.41	3.82	
Ser -5	117.7	8.23	4.35	3.78	
Leu -4	125.0	8.02	4.29	1.59	$^{\gamma}CH$ 1.27; $^{\delta}CH_3$ 0.79
Phe -3	120.1	8.08	5.00	3.43, 3.23	$^{\delta}H$ 7.27; $^{\varepsilon}H$ n.o.; $^{\zeta}H$ n.o.
Val -2	120.7	7.37	3.98	2.18	$^{\gamma}CH_3$ 0.86
Asn -1	121.8	8.47	4.37	2.90, 2.81	$^{\gamma}NH_2$ n.o.
Gly 1	108.9	8.06	n.o.		
Pro 2			4.42	2.10	$^{\gamma}CH_2$ 1.85; $^{\delta}CH_2$ 3.45
Glu 3	118.3	8.25	4.66	2.12, 1.92	$^{\gamma}CH_2$ 2.41, 2.26
Thr 4	116.1	7.85	4.32	3.78	$^{\gamma}CH_3$ 1.31
Leu 5	119.2	7.57	3.95	2.03	$^{\gamma}CH$ 1.66; $^{\delta}CH_3$ 0.87
Cys 6	118.3	8.33	4.89	2.89, 2.61	
Gly 7	108.28	7.63	3.91, 3.58		
Ala 8	128.15	8.83	3.86	1.39	
Glu 9	115.1	7.94	4.01	2.17, 2.01	$^{\gamma}CH_2$ 2.63, 2.51
Leu 10	120.3	6.87	4.09	1.90, 1.81	$^{\gamma}CH$ 1.2.1; $^{\delta}CH_3$ 1.14, 0.88
Val 11	117.9	7.38	3.26	1.97	$^{\gamma}CH_3$ 0.99, 0.82
Asp 12	117.2	8.11	4.03	2.85, 2.74	
Ala 13	122.2	7.76	3.99	1.35	
Leu 14	118.5	8.24	3.73	1.45	$^{\gamma}CH$ 1.91; $^{\delta}CH_3$ 0.86, 0.73
Gln 15	118.5	8.12	4.04	2.23, 2.00	$^{\gamma}CH_2$ 2.52, 2.37, $^{\delta}NH_2$ 7.11, 6.69, $^{\varepsilon}N$ 83.2
Phe 16	117.9	7.38	4.35	3.16	$^{\delta}H$ 7.23; $^{\varepsilon}H$ n.o.; $^{\zeta}H$ n.o.

Val 17	117.7	8.60	3.52	1.94	$^{\gamma}CH_3$ 1.02, 0.80
Cys 18	115.1	8.69	4.71	3.25, 2.86	
Gly 19	108.1	7.41	4.02, 3.92		
Asp 20	122.5	8.77	4.48	2.84	
Arg 21	118.3	8.00	4.10	1.94	$^{\gamma}CH_2$ 1.77; $^{\delta}CH_2$ 3.22; $^{\varepsilon}NH$ 7.03, $^{\varepsilon}N$ 90.8
Gly 22	104.0	7.33	3.93, 3.69		
Phe 23	114.2	7.46	4.94	3.20, 2.80	$^{\delta}H$ 6.88; $^{\varepsilon}H$ 7.16; $^{\zeta}H$ n.o.
Tyr 24	119.6	8.41	4.49	3.18, 2.88	$^{\delta}H$ 6.77, $^{\varepsilon}H$ 6.94
Phe 25	120.3	7.86	4.50	2.97, 2.85	$^{\delta}H$ 7.17; $^{\varepsilon}H$ 6.87; $^{\zeta}H$ n.o.
Asn 26	120.1	7.88	4.49	2.56	$^{\gamma}NH_2$ 7.50, 6.80, $^{\delta}N$ 112.9
Lys 27	121.8	7.99	4.40	1.72, 1.62	$^{\gamma}CH_2$ 1.34; $^{\delta}CH_2$ 2.59
Pro 28			4.43	2.18, 2.11	$^{\gamma}CH_2$ 1.86; $^{\delta}CH_2$ 3.41
Thr 29	113.1	8.07	4.27	4.15	$^{\gamma}CH_3$ 1.14
Gly 30	110.2	8.20	3.90, 3.80		
Tyr 31	120.1	8.04	4.47	2.97, 2.86	$^{\delta}H$ 7.03; $^{\varepsilon}H$ 6.88
Gly 32	110.7	8.30	3.93, 3.79		
Ser 33	115.5	8.13	4.38	3.91, 3.82	
Ser 34	117.0	8.11	4.39	3.82	
Ser 35	117.2	8.32	4.39	3.85	
Arg 36	122.0	8.08	4.25	1.80, 1.65	$^{\gamma}CH_2$ 1.55; $^{\delta}CH_2$ 3.11; $^{\varepsilon}NH$ 7.08, $^{\varepsilon}N$ 85.43
Arg 37	121.2	8.08	4.27	1.79, 1.67	$^{\gamma}CH_2$ 1.54; $^{\delta}CH_2$ 3.07; $^{\varepsilon}NH$ 7.17, $^{\varepsilon}N$ 85.43.
Ala 38	126.0	8.16	4.51	1.32	
Pro 39			4.36	2.16, 2.06	$^{\gamma}CH_2$ 1.95, 1.86; $^{\delta}CH_2$ 3.40
Gln 40	118.3	8.46	4.15	1.95, 1.82	$^{\gamma}CH_2$ 2.29; $^{\delta}NH_2$ 7.38, 6.76, $^{\varepsilon}N$ 111.91
Thr 41	114.2	7.85	4.26	4.04	$^{\gamma}CH_3$ 1.38, 1.18
Gly 42	109.4	8.26	4.06, 3.95		
Ile 43	121.4	7.74	3.85	2.00	$^{\gamma}CH_2$ 1.84, 1.58; $^{\gamma}CH_3$ 1.14; $^{\delta}CH_3$ 0.87
Val 44	120.3	7.79	4.42	2.03	$^{\gamma}CH_3$ 0.89
Asp 45	118.1	8.06	4.30	2.83	
Glu 46	118.1	7.83	4.41	1.76	$^{\gamma}CH_2$ 2.15
Cys 47	115.5	7.22	4.93	3.15, 2.73	
Cys 48	115.7	8.43	4.40	3.42, 3.24	
Phe 49	113.7	7.89	4.73	3.26, 2.85	$^{\delta}H$ 7.32; $^{\varepsilon}H$ n.o.; $^{\zeta}H$ n.o.
Arg 50	119.4	7.95	4.53	1.89	$^{\gamma}CH_2$ 1.57; $^{\delta}CH_2$ n.o.; $^{\varepsilon}NH$ n.o.

Ser 51	112.9	7.81	4.16	3.69	
Cys 52	119.0	8.21	4.71	2.99, 2.85	
Asp 53	116.1	7.59	4.24	3.36, 2.95	
Leu 54	122.5	7.57	3.95	1.30	$^{\gamma}$CH 1.72; $^{\delta}CH_3$ 0.94
Arg 55	119.4	7.92	4.10	1.90	$^{\gamma}CH_2$ 1.79; $^{\delta}CH_2$ 3.28; $^{\varepsilon}$NH n.o.
Arg 56	116.8	7.86	3.85	1.74	$^{\gamma}CH_2$ 1.59; $^{\delta}CH_2$ 3.15; $^{\varepsilon}$NH 7.37
Leu 57	117.0	7.73	3.88	1.93	$^{\gamma}$CH 1.81; $^{\delta}CH_3$ 1.03, 0.77
Glu 58	113.7	8.15	4.05	2.07, 1.95	$^{\gamma}CH_2$ 2.49, 2.29
Met 59	117.9	7.53	4.18	2.32, 2.12	$^{\gamma}CH_2$ 2.89, 2.59; $^{\varepsilon}$CH 1.82
Leu 60	117.9	7.76	4.03	1.78	$^{\gamma}$CH 2.15, $^{\delta}CH_3$ 0.89
Cys 61	115.3	7.22	4.93	3.16, 2.73	
Ala 62	126.4	8.20	4.33	1.77	
Pro 63			4.33	2.15, 2.05	$^{\gamma}CH_2$ 1.85, 1.79; $^{\delta}CH_2$ 3.39
Leu 64	123.6	8.24	4.53	1.73, 1.65	$^{\gamma}$CH 1.49; $^{\delta}CH_3$ 0.85
Lys 65	122.0	8.18	4.14	1.49	$^{\gamma}CH_2$ 1.95, 1.82; $^{\delta}CH_2$ 2.27
Pro 66			4.37	2.16, 2.10	$^{\gamma}CH_2$ 1.88; $^{\delta}CH_2$ 3.39
Ala 67	124.6	8.30	4.32	1.33	
Lys 68	122.3	8.19	4.03	1.54, 1.41	$^{\gamma}CH_2$ 1.83; $^{\delta}CH_2$ n.o.; $^{\varepsilon}CH_2$ n.o.
Ser 69	112.9	7.93	4.14	3.88, 3.77	
Ala 70	128.6	8.09	4.17	1.31	

APPENDIX 4

Dihedral Angle and Distance Restraints

Distance and dihedral angle restraints used in the three dimensional structure calculation of Long-[Arg^3]-IGF-I

Table 1: Dihedral angle restraints

Residue	Dihedral Angle	Lower Bound	Upper Bound
1 Met	phi	-100	-20
4 Ala	phi	-100	-20
5 Met	phi	-100	-20
7 Leu	phi	-160	-80
7 Leu	psi	60	180
8 Ser	phi	-160	-80
8 Ser	psi	60	180
9 Ser	phi	-160	-80
9 Ser	psi	60	180
12 Val	phi	-100	-20
13 Asn	phi	-100	-20
16 Arg	phi	-100	-20
17 Thr	phi	-100	-20
18 Leu	phi	-100	-20
19 Cys	phi	-100	-20
21 Ala	phi	-90	-30
21 Ala	psi	-80	-20
22 Glu	phi	-90	-30
22 Glu	psi	-80	-20
23 Leu	phi	-90	-30
23 Leu	psi	-80	-20
24 Val	phi	-90	-30
24 Val	psi	-80	-20
26 Ala	phi	-90	-30
26 Ala	psi	-80	-20
28 Gln	phi	-90	-30
28 Gln	psi	-80	-20
29 Phe	phi	-90	-30
29 Phe	psi	-80	-20
30 Val	phi	-90	-30

30 Val	psi	-80	-20
31 Cys	phi	-90	-30
31 Cys	psi	-80	-20
33 Asp	phi	-90	-30
36 Phe	phi	-100	-20
37 Tyr	phi	-100	-20
39 Asn	phi	-100	-40
40 Lys	phi	-100	-40
42 Thr	phi	-160	-40
43 Gly	phi	-100	180
44 Tyr	phi	-160	-40
44 Tyr	psi	60	-80
45 Gly	phi	-160	-80
46 Ser	phi	-160	-80
46 Ser	psi	60	180
47 Ser	phi	-160	-80
47 Ser	psi	60	180
48 Ser	phi	-160	-80
48 Ser	psi	60	180
49 Arg	phi	-160	-40
50 Arg	phi	-160	-40
51 Ala	phi	-160	-40
53 Gln	phi	-100	-20
54 Thr	phi	-90	-30
55 Gly	phi	-100	-20
56 Ile	phi	-90	-30
56 Ile	psi	-100	-20
57 Val	phi	-100	-20
58 Asp	phi	-90	-30
58 Asp	psi	-80	-20
59 Glu	phi	-90	-30
59 Glu	psi	-80	-20
61 Cys	phi	-90	-30
61 Cys	psi	-80	-20
62 Phe	phi	-90	-30

62 Phe	psi	-80	-20
63 Arg	phi	-100	-20
64 Ser	phi	-100	-20
65 Cys	phi	-160	-80
65 Cys	psi	60	180
66 Asp	phi	-160	-40
67 Leu	phi	-90	-30
67 Leu	psi	-80	-20
68 Arg	phi	-90	-30
68 Arg	psi	-80	-20
69 Arg	phi	-90	-30
69 Arg	psi	-80	-20
71 Glu	phi	-90	-30
71 Glu	psi	-80	-20
72 Met	phi	-100	-20
72 Met	psi	-80	-20
73 Tyr	phi	-100	-20
73 Tyr	psi	-80	-20
74 Cys	phi	-100	-20
75 Ala	phi	-100	-20
77 Leu	phi	-160	-40
78 Lys	phi	-160	-40
80 Ala	phi	-160	-80
82 Ser	phi	-100	-20
83 Ala	phi	-100	-20

Table 2: Lower Bounds Distance Restraints

Residue 1	Atom	Residue 2	Atom	Lower Bound
19 Cys	CB	61 Cys	SG	3.00
	SG	61 Cys	CB	3.00
	SG	61 Cys	SG	2.00
21 Ala	O	25 Asp	HN	1.50
	O	25 Asp	N	2.30
22 Glu	O	26 Ala	HN	1.50
	O	26 Ala	N	2.30
23 Leu	O	27 Leu	HN	1.50
	O	27 Leu	N	2.30

24 Val	O	28 Gln	HN	1.50
	O	28 Gln	N	2.30
25 Asp	O	29 Phe	HN	1.50
	O	29 Phe	N	2.30
26 Ala	O	30 Val	HN	1.50
	O	30 Val	N	2.30
27 Leu	O	31 Cys	HN	1.50
	O	31 Cys	N	2.30
31 Cys	CB	74 Cys	SG	3.00
	SG	74 Cys	CB	3.00
	SG	74 Cys	SG	2.00
32 Gly	O	35 Gly	HN	1.50
	O	35 Gly	N	2.30
57 Val	O	61 Cys	HN	1.50
	O	61 Cys	N	2.30
58 Asp	O	62 Phe	HN	1.50
	O	62 Phe	N	2.30
59 Glu	O	63 Arg	HN	1.50
	O	63 Arg	N	2.30
60 Cys	CB	65 Cys	SG	3.00
	SG	65 Cys	CB	3.00
	SG	65 Cys	SG	2.00
67 Leu	O	71 Glu	HN	1.50
	O	71 Glu	N	2.30
68 Arg	O	72 Met	HN	1.50
	O	72 Met	N	2.30

Table 3: Upper Bound Distance Restraints

Residue 1	Atom	Residue 2	Atom	Upper Bound	Weighting
1 Met	HB2	8 Ser	HN	6.50	1.00
	HB3	8 Ser	HN	6.50	1.00
2 Phe	HN	2 Phe	HA	2.40	5.00
	QB	5 Met	HN	3.99	4.00
3 Pro	HA	4 Ala	HN	4.50	3.00
4 Ala	HN	4 Ala	HA	2.40	5.00
5 Met	HN	5 Met	QG	6.38	5.00
7 Leu	QD1	8 Ser	HN	6.00	2.00
	QD2	8 Ser	HN	6.00	2.00
8 Ser	HN	9 Ser	HN	2.52	5.00
	QB	12 Val	HN	5.45	2.00
9 Ser	HN	9 Ser	HA	2.90	5.00
	HN	11 Phe	QB	6.00	2.00
10 Leu	HN	12 Val	HN	3.30	5.00
11 Phe	HN	11 Phe	HB2	3.30	5.00
	HN	11 Phe	HB3	3.30	5.00
	HN	12 Val	HN	3.02	6.00
	HN	12 Val	HA	3.83	6.00
	HN	12 Val	HB	3.48	6.00
	HN	12 Val	QQG	5.89	5.00
	HA	19 Cys	HN	6.50	4.00
	HA	20 Gly	HN	6.50	4.00

12 Val	HN	12 Val	HB	3.02	5.00
	HN	13 Asn	HN	4.17	6.00
	HN	19 Cys	HN	6.50	4.00
	HA	14 Gly	HN	4.00	5.00
	HA	19 Cys	HN	6.50	4.00
	HB	19 Cys	HN	6.50	4.00
	QQG	20 Gly	HN	6.50	4.00
	QQG	19 Cys	HN	6.50	4.00
13 Asn	HN	13 Asn	HB2	3.76	5.00
	HN	13 Asn	HB3	3.76	5.00
	HN	14 Gly	HN	4.63	6.00
16 Arg	HN	16 Arg	HB2	3.48	5.00
	HN	16 Arg	HB3	3.48	5.00
	HN	16 Arg	QB	3.13	5.00
	HN	16 Arg	HG2	4.48	5.00
	HN	16 Arg	HG3	4.48	5.00
	HN	16 Arg	QG	4.50	5.00
	HN	16 Arg	QD	5.82	4.00
	HN	16 Arg	NH1	4.69	4.00
	HN	16 Arg	NH2	4.54	4.00
	HN	17 Thr	HN	3.55	6.00
	HN	21 Ala	HN	6.50	
	HA	17 Thr	HN	5.00	3.00
	HG2	17 Thr	HN	6.00	3.00
17 Thr	HN	18 Leu	HN	2.71	6.00
18 Leu	HN	18 Leu	HG	5.38	3.00
	HN	18 Leu	QQD	7.63	3.00
19 Cys	HN	19 Cys	HA	2.68	5.00
	HN	19 Cys	QB	3.76	5.00
	HN	20 Gly	HN	2.99	6.00
	HA	20 Gly	HN	2.93	7.00
	CB	61 Cys	SG	3.10	8.00
	HB2	20 Gly	HN	4.14	6.00
	HB3	20 Gly	HN	4.14	6.00
	QB	20 Gly	HN	4.50	6.00
	SG	61 Cys	CB	3.10	8.00
	SG	61 Cys	SG	2.10	8.00
20 Gly	HA1	21 Ala	HN	3.48	5.00
	HA2	21 Ala	HN	3.48	5.00
	QA	22 Glu	HN	5.07	6.00
21 Ala	HN	22 Glu	HN	3.33	6.00
	HA	22 Glu	HN	3.27	6.00
	HA	24 Val	HN	3.80	8.00
	HA	25 Asp	HN	4.50	7.00
	QB	22 Glu	HN	4.23	6.00
	QB	23 Leu	HN	6.53	3.00
	QB	24 Val	HN	6.00	4.00
	O	25 Asp	HN	2.30	8.00
	O	25 Asp	N	3.30	8.00

22 Glu	HN	22 Glu	HB2	3.08	5.00
	HN	22 Glu	HB3	3.08	5.00
	HN	22 Glu	QB	3.08	5.00
	HN	22 Glu	HG2	4.97	5.00
	HN	22 Glu	HG3	4.97	5.00
	HN	22 Glu	QG	5.00	5.00
	HN	23 Leu	HN	3.17	6.00
	HN	24 Val	HN	4.50	6.00
	HA	24 Val	HN	3.61	6.00
	HA	25 Asp	HN	3.80	8.00
	HA	26 Ala	HN	4.50	7.00
	O	26 Ala	HN	2.30	8.00
	O	26 Ala	N	2.30	8.00
23 Leu	HN	23 Leu	HG	4.88	4.00
	HN	23 Leu	QD1	6.00	3.00
	HN	23 Leu	QD2	6.00	3.00
	HN	23 Leu	QQD	6.00	3.00
	HN	24 Val	HN	3.64	6.00
	HA	26 Ala	HN	3.80	8.00
	HA	27 Leu	HN	4.50	7.00
	QB	24 Val	HN	4.92	5.00
	HA	36 Phe	HN	7.00	
	QD2	27 Leu	HN	6.50	5.00
	O	27 Leu	HN	2.30	8.00
	O	27 Leu	N	3.30	8.00
24 Val	HN	24 Val	HB	3.08	5.00
	HA	27 Leu	HN	3.50	8.00
	HA	28 Gln	HN	4.40	7.00
	HN	25 Asp	HN	3.00	6.00
	HN	26 Ala	HN	4.40	6.00
	O	28 Gln	HN	2.30	8.00
	O	28 Gln	N	3.30	8.00
	HB	56 Ile	QG2	6.00	4.00
	HB	57 Val	QG2	6.00	4.00
	HB	57 Val	QG1	6.00	4.00
	QG1	38 Phe	QD	6.00	4.00
	QG2	38 Phe	QD	6.00	4.00
	QG1	38 Phe	QE	6.00	4.00
	QG2	38 Phe	QE	6.00	4.00
25 Asp	HB2	26 Ala	HN	3.80	6.00
	HB3	26 Ala	HN	3.80	6.00
	QB	26 Ala	HN	3.80	6.00
	HN	26 Ala	HN	3.30	6.00
	HA	28 Gln	HN	3.80	8.00
	HA	29 Phe	HN	4.50	7.00
	O	29 Phe	HN	2.30	8.00
	O	29 Phe	N	3.30	8.00
26 Ala	HN	27 Leu	HN	2.99	6.00
	HN	28 Gln	HN	3.72	6.00
	HN	67 Leu	QB	7.00	
	HA	29 Phe	HN	3.80	8.00
	HA	30 Val	HN	4.50	6.00
	QB	28 Gln	HN	6.53	5.00
	O	30 Val	HN	2.30	8.00
	O	30 Val	N	3.30	8.00
	QB	56 Ile	QG1	6.00	4.00
	QB	56 Ile	QG2	6.00	4.00

27 Leu	HN	27 Leu	HB2	3.50	4.00
	HN	27 Leu	HB3	3.50	4.00
	HN	27 Leu	QB	3.40	4.00
	HN	27 Leu	HG	5.47	4.00
	HN	27 Leu	QD1	6.00	3.00
	HN	27 Leu	QD2	6.00	3.00
	HN	27 Leu	QQD	6.00	3.00
	HN	28 Gln	HN	2.83	6.00
	HA	30 Val	HN	3.80	8.00
	HA	31 Cys	HN	4.50	7.00
	QB	28 Gln	HN	4.30	6.00
	QB	29 Phe	HN	5.14	5.00
	QB	73 Tyr	HN	7.49	3.00
	HG	28 Gln	HN	4.32	5.00
	HG	73 Tyr	HN	5.38	3.00
	QQD	28 Gln	HN	6.92	4.00
	QQD	73 Tyr	HN	8.10	3.00
	O	31 Cys	HN	2.30	8.00
	O	31 Cys	N	3.30	8.00
	QB	36 Phe	QD	6.00	4.00
	QD1	36 Phe	QD	6.00	4.00
	QD2	36 Phe	QD	6.00	4.00
	QB	36 Phe	QE	6.00	4.00
	QD1	36 Phe	QE	6.00	4.00
	QD2	36 Phe	QE	6.00	4.00
	QD2	56 Ile	HB	6.50	3.00
	QD1	56 Ile	HA	6.50	2.00
	QD2	56 Ile	QG1	6.50	2.00
	QD2	56 Ile	QG2	6.50	3.00
	QD1	56 Ile	QD1	6.50	3.00
	QD2	70 Leu	GB	6.50	2.00
	QD2	70 Leu	QD1	6.50	3.00
	QD2	70 Leu	QD2	6.50	3.00
	QD1	57 Val	QG1	6.50	3.00
	QD1	57 Val	QG2	6.50	3.00
28 Gln	HN	28 Gln	HB2	3.02	5.00
	HN	28 Gln	HB3	3.02	5.00
	HN	28 Gln	HG2	5.07	5.00
	HN	28 Gln	HG3	5.07	5.00
	HN	28 Gln	QG	5.00	5.00
	HN	28 Gln	QE2	6.37	4.00
	HN	29 Phe	HN	3.17	6.00
	HA	29 Phe	HN	3.05	8.00
	HA	31 Cys	HN	4.11	8.00
	HB2	29 Phe	HN	3.83	6.00
	HB3	29 Phe	HN	3.83	6.00
	QB	29 Phe	HN	4.00	6.00

29 Phe	HN	30 Val	HN	3.12	6.00
	HA	32 Gly	HN	4.23	6.00
	QD	30 Val	HN	7.62	3.00
	HA	30 Val	HN	3.50	8.00
	QD	67 Leu	QD1	6.50	3.00
	QD	67 Leu	HG	6.50	3.00
	QD	67 Leu	QD2	6.50	3.00
	QE	67 Leu	QD1	6.50	3.00
	QE	67 Leu	HG	6.50	3.00
	QE	67 Leu	QD2	6.50	3.00
	QD	70 Leu	QD1	6.50	3.00
	QE	70 Leu	QD2	6.50	3.00
30 Val	HN	30 Val	HB	3.05	5.00
	HN	30 Val	QG1	4.79	5.00
	HN	30 Val	QG2	4.79	5.00
	HN	31 Cys	HN	3.20	6.00
	HA	31 Cys	HN	3.50	8.00
	HB	31 Cys	HN	5.00	6.00
	HB	32 Gly	HN	5.00	5.00
	QG1	31 Cys	HN	5.13	4.00
	QG2	31 Cys	HN	5.13	3.00
	QQG	31 Cys	HN	4.76	3.00
31 Cys	HN	31 Cys	HB2	3.05	5.00
	HN	31 Cys	HB3	3.05	5.00
	HN	32 Gly	HN	2.90	6.00
	HN	32 Gly	QA	5.23	6.00
	HA	32 Gly	HN	3.42	6.00
	CB	74 Cys	SG	3.10	8.00
	HB2	32 Gly	HN	3.52	6.00
	HB3	32 Gly	HN	3.52	6.00
	QB	32 Gly	HN	3.26	6.00
	QB	35 Gly	HN	6.50	
	SG	74 Cys	CB	3.10	8.00
	SG	74 Cys	SG	2.10	8.00
32 Gly	HN	33 Asp	HN	3.36	6.00
	QA	34 Arg	HN	4.83	6.00
	O	35 Gly	HN	2.30	8.00
	O	35 Gly	N	3.30	8.00
33 Asp	HN	34 Arg	HN	3.08	6.00
	HN	35 Gly	HN	4.91	6.00
	HA	34 Arg	HN	4.50	6.00
	HA	35 Gly	HN	3.36	6.00
	QB	34 Arg	HN	4.42	6.00
34 Arg	HN	34 Arg	HG2	3.33	5.00
	HN	34 Arg	HG3	3.33	5.00
	HN	34 Arg	QD	5.35	4.00
	HN	34 Arg	NH1	6.00	4.00
	HN	35 Gly	HN	2.99	6.00
	HN	35 Gly	QA	5.57	6.00
	HG2	35 Gly	HN	5.50	6.00
	HG3	35 Gly	HN	5.50	6.00
	QG	35 Gly	HN	5.18	3.00
	QD	35 Gly	HN	6.38	2.00
35 Gly	HN	36 Phe	HN	5.07	6.00
	HA1	36 Phe	HN	3.11	6.00
	HA2	36 Phe	HN	3.11	6.00

36 Phe	HN	36 Phe	HB2	3.95	5.00
	HN	36 Phe	HB3	3.95	5.00
	HN	36 Phe	QB	3.77	5.00
	HN	37 Tyr	HN	5.50	6.00
	HA	37 Tyr	HN	3.02	6.00
	QD	75 Ala	HA	6.50	3.00
	QD	75 Ala	QB	6.50	3.00
	QE	75 Ala	HA	6.50	3.00
	QE	75 Ala	QB	6.50	3.00
37 Tyr	HN	37 Tyr	HB2	3.21	5.00
	HN	37 Tyr	HB3	3.21	5.00
	HN	37 Tyr	QE	7.63	3.00
38 Phe	HN	38 Phe	HB2	3.64	5.00
	HN	38 Phe	HB3	3.64	5.00
	HN	38 Phe	QB	3.36	5.00
	HN	39 Asn	HN	3.50	6.00
	HA	39 Asn	HN	3.21	6.00
	HB2	39 Asn	HN	3.52	6.00
	HB3	39 Asn	HN	3.52	6.00
	QB	39 Asn	HN	3.34	6.00
	QE	56 Ile	QG1	6.00	3.00
	QD	56 Ile	QG1	6.00	3.00
	QE	56 Ile	QG2	6.50	3.00
	QD	56 Ile	QG2	6.50	3.00
39 Asn	HN	39 Asn	HB2	3.30	5.00
	HN	39 Asn	HB3	3.30	5.00
	HA	40 Lys	HN	3.50	6.00
	QB	40 Lys	HN	4.52	6.00
	HA	73 Tyr	QD	6.00	3.00
	HA	73 Tyr	QE	6.00	3.00
40 Lys	HN	40 Lys	QG	6.38	3.00
	HN	40 Lys	QE	6.38	3.00
	HN	42 Thr	HN	5.00	4.00
41 Pro	HA	42 Thr	HN	4.50	3.00
42 Thr	HN	42 Thr	HA	2.46	5.00
	HN	42 Thr	HB	2.96	5.00
	HN	42 Thr	QG2	4.54	5.00
	HB	43 Gly	HN	3.36	6.00
	HA	43 Gly	HN	4.50	6.00
43 Gly	HN	43 Gly	HA1	2.43	5.00
	HN	43 Gly	HA2	2.43	5.00
	QA	44 Tyr	HN	3.50	6.00
44 Tyr	HN	44 Tyr	HB2	2.90	5.00
	HN	44 Tyr	HB3	2.90	5.00
	HN	45 Gly	HN	4.00	6.00
	HA	45 Gly	HN	3.00	6.00
	HB2	45 Gly	HN	4.50	6.00
	HB3	45 Gly	HN	4.50	6.00
	QB	45 Gly	HN	4.00	6.00
	HB2	46 Ser	HN	5.50	3.00
	HB3	46 Ser	HN	5.50	3.00
	HB2	47 Ser	HN	6.50	1.00
45 Gly	HN	45 Gly	HA1	2.59	5.00
	HN	45 Gly	HA2	2.59	5.00
	HN	46 Ser	HN	4.00	6.00
	HA1	46 ser	HN	3.50	6.00
	HA2	46 Ser	HN	3.50	6.00

46 Ser	HN	46 Ser	HB2	2.56	5.00
	HN	46 Ser	HB3	2.56	5.00
	HN	47 Ser	HN	3.76	6.00
	HA	47 Ser	HN	2.50	6.00
	HB2	47 Ser	HN	5.00	3.00
	HB3	47 Ser	HN	5.00	3.00
47 Ser	HN	47 Ser	HA	2.56	5.00
	HN	48 Ser	HN	3.86	6.00
	HA	48 Ser	HN	2.50	6.00
48 Ser	HA	49 Arg	HN	2.50	6.00
	HN	49 Arg	HN	4.40	6.00
49 Arg	HN	49 Arg	HA	2.56	5.00
	HN	49 Arg	HB2	4.00	5.00
	HN	49 Arg	HB3	4.00	5.00
	HN	49 Arg	QB	4.50	5.00
	HN	49 Arg	QG	4.73	4.00
	HN	49 Arg	NH1	5.50	3.00
	HA	50 Arg	HN	2.50	6.00
	HN	50 Arg	HN	4.40	6.00
50 Arg	HN	50 Arg	HA	2.40	5.00
	HN	50 Arg	HG2	3.73	5.00
	HN	50 Arg	HG3	3.73	5.00
	HN	50 Arg	QG	3.44	5.00
	HA	53 Gln	HN	6.50	2.00
	HA	51 Ala	HN	2.50	6.00
	HN	51 Ala	HN	4.40	6.00
	HB2	51 Ala	HN	5.00	6.00
	HB3	51 Ala	HN	5.00	6.00
	HG2	51 Ala	HN	6.00	2.00
53 Gln	HN	53 Gln	HA	2.40	5.00
	HN	53 Gln	HB2	3.30	5.00
	HN	53 Gln	HB3	3.30	5.00
	HN	53 Gln	QB	3.50	5.00
	HN	53 Gln	HG2	4.50	4.00
	HN	53 Gln	HG3	4.50	4.00
	HN	53 Gln	QG	5.00	4.00
	HN	54 Thr	HN	2.99	6.00
	HG2	54 Thr	HN	3.95	6.00
	HG3	54 Thr	HN	3.95	6.00
	QG	54 Thr	HN	3.64	6.00
	HA	54 Thr	HN	4.50	6.00
54 Thr	HN	54 Thr	HA	2.40	5.00
	HN	54 Thr	HB	3.05	5.00
	HN	54 Thr	QG2	4.17	4.00
	HN	55 Gly	HN	3.08	6.00
	HA	55 Gly	HN	2.96	6.00
	QG2	55 Gly	HN	6.12	3.00
	HN	56 Ile	HN	5.00	3.00
55 Gly	HN	56 Ile	HN	3.89	6.00
	HA1	56 Ile	HN	2.99	6.00
	HA2	56 Ile	HN	2.99	6.00
	QA	56 Ile	HN	2.79	6.00

56 Ile	HN	56 Ile	HB	3.24	5.00
	HN	56 Ile	QG2	4.83	5.00
	N	56 Ile	QG1	4.08	5.00
	HN	57 Val	HN	3.05	6.00
	HN	73 Tyr	QD	7.24	5.00
	HA	59 Glu	HN	3.80	8.00
	HA	60 Cys	HN	4.50	7.00
	HB	57 Val	HN	4.51	6.00
	QG1	73 Tyr	QD	6.00	3.00
	HB	73 Tyr	QD	6.00	3.00
57 Val	HN	57 Val	HB	3.11	5.00
	HN	57 Val	QG1	4.33	5.00
	HN	57 Val	QG2	4.33	5.00
	HN	57 Val	QQG	4.50	4.00
	HN	58 Asp	HN	3.21	6.00
	HN	60 Cys	HN	3.21	7.00
	HA	58 Asp	HN	3.48	6.00
	HA	60 Cys	HN	4.00	8.00
	HA	61 Cys	HN	4.50	7.00
	HB	58 Asp	HN	3.33	6.00
	QG1	58 Asp	HN	6.06	4.00
	QG2	58 Asp	HN	6.06	4.00
	QQG	58 Asp	HN	5.31	4.00
	QQG	60 Cys	HN	7.50	4.00
	O	61 Cys	HN	2.30	8.00
	O	61 Cys	N	3.30	8.00
58 Asp	HA	59 Glu	HN	3.50	7.00
	HN	59 Glu	HN	3.20	6.00
	HA	60 Cys	HN	3.80	8.00
	HA	62 Phe	HN	4.50	7.00
	O	62 Phe	HN	2.30	8.00
	O	62 Phe	N	3.30	8.00
59 Glu	HN	59 Glu	HA	2.80	5.00
	HN	59 Glu	QG	4.27	5.00
	HN	60 Cys	HN	3.27	4.00
	HA	62 Phe	HN	3.80	8.00
	HA	63 Arg	HN	4.50	7.00
	O	63 Arg	HN	2.30	8.00
	O	63 Arg	N	3.30	8.00
60 Cys	HN	60 Cys	HB2	3.48	5.00
	HN	60 Cys	HB3	3.48	5.00
	HN	60 Cys	QB	3.50	5.00
	HN	62 Phe	HN	3.42	5.00
	HA	61 Cys	HN	3.58	6.00
	HN	61 Cys	HN	3.20	6.00
	HA	63 Arg	HN	3.80	8.00
	CB	65 Cys	SG	3.10	8.00
	SG	65 Cys	CB	3.10	8.00
	SG	65 Cys	SG	2.10	8.00
61 Cys	HN	61 Cys	HB2	3.70	5.00
	HN	61 Cys	HB3	3.70	5.00
	HN	61 Cys	QB	3.80	5.00
	HN	62 Phe	HN	3.01	6.00
	HN	63 Arg	HN	4.04	6.00
	HB2	62 Phe	HN	3.39	6.00
	HB3	62 Phe	HN	3.39	6.00
	QB	62 Phe	HN	3.20	6.00

62 Phe	HN	62 Phe	HB2	3.64	5.00
	HN	62 Phe	HB3	3.64	5.00
	HN	62 Phe	QB	3.42	5.00
	HN	63 Arg	HN	2.96	6.00
	HN	64 Ser	HN	6.17	4.00
	HN	65 Cys	HN	6.50	2.00
	HA	64 Ser	HN	6.01	2.00
	HB2	63 Arg	HN	4.32	6.00
	HB3	63 Arg	HN	4.32	6.00
	QB	63 Arg	HN	4.08	6.00
63 Arg	HN	63 Arg	QG	5.73	5.00
	HN	63 Arg	QD	6.19	5.00
	HN	64 Ser	HN	4.00	6.00
	HN	65 Cys	HN	5.61	2.00
	HA	64 Ser	HN	2.71	6.00
	QB	64 Ser	HN	4.90	4.00
	QG	64 Ser	HN	5.79	3.00
64 Ser	HA	66 Asp	HN	4.90	2.00
	QB	66 Asp	HN	5.64	2.00
65 Cys	HN	65 Cys	HB2	3.76	5.00
	HN	65 Cys	HB3	3.76	5.00
	HN	65 Cys	QB	3.34	5.00
66 Asp	HN	66 Asp	HB2	3.73	5.00
	HN	66 Asp	HB3	3.73	5.00
	HN	66 Asp	QB	3.28	5.00
	HN	67 Leu	HN	4.00	6.00
67 Leu	HN	67 Leu	QD1	5.69	4.00
	HN	67 Leu	QD2	5.69	4.00
	HN	67 Leu	QQD	4.98	4.00
	HN	68 Arg	HN	3.00	6.00
	HA	69 Arg	HN	3.58	6.00
	HA	70 Leu	HN	3.80	8.00
	HA	71 Glu	HN	4.58	7.00
	QQD	68 Arg	HN	6.67	3.00
	QQD	69 Arg	HN	7.85	3.00
	O	71 Glu	HN	2.30	8.00
	O	71 Glu	N	3.30	8.00
68 Arg	HN	68 Arg	HB2	3.59	5.00
	HN	68 Arg	HB3	3.59	5.00
	HN	68 Arg	HG2	5.00	4.00
	HN	68 Arg	HG3	5.00	4.00
	HN	68 Arg	QG	5.00	4.00
	HN	68 Arg	QD	6.38	3.00
	HA	69 Arg	HN	3.52	6.00
	HA	71 Glu	HN	3.80	8.00
	HA	72 Met	HN	4.50	7.00
	QB	69 Arg	HN	4.05	6.00
	HN	69 Arg	HN	3.20	6.00
	O	72 Met	HN	2.30	8.00
	O	72 Met	N	3.30	8.00

69 Arg	HN	69 Arg	HB2	3.24	5.00
	HN	69 Arg	HB3	3.24	5.00
	HN	69 Arg	QB	2.81	5.00
	HN	69 Arg	HG2	4.48	4.00
	HN	69 Arg	HG3	4.48	4.00
	HN	69 Arg	QG	4.50	4.00
	HN	69 Arg	QD	5.79	3.00
	HN	70 Leu	HN	3.10	7.00
	HA	71 Glu	HN	3.89	6.00
	HA	73 Tyr	HN	4.50	7.00
	HA	72 Met	HN	3.80	8.00
70 Leu	HN	70 Leu	HA	2.40	5.00
	HN	70 Leu	HB2	3.50	5.00
	HN	70 Leu	HB3	3.50	5.00
	HN	70 Leu	QB	3.50	5.00
	HN	70 Leu	QQD	7.32	3.00
	HA	73 Tyr	HN	3.80	8.00
	QB	71 Glu	HN	5.17	6.00
	QQD	71 Glu	HN	7.60	3.00
	HN	71 Glu	HN	3.20	3.00
	QQD	72 Met	HN	7.63	2.00
71 Glu	HN	71 Glu	HB2	3.67	5.00
	HN	71 Glu	HB3	3.67	5.00
	HN	71 Glu	QB	3.35	5.00
	HN	71 Glu	HG2	4.00	5.00
	HN	71 Glu	HG3	4.00	5.00
	HN	72 Met	HN	3.01	7.00
72 Met	HN	72 Met	HA	2.56	5.00
	HN	72 Met	HB2	3.30	5.00
	HN	72 Met	HB3	3.30	5.00
	HN	72 Met	HG2	5.50	5.00
	HN	72 Met	HG3	5.50	5.00
	HN	72 Met	QG	5.50	5.00
	HN	72 Met	QE	6.53	3.00
	HN	74 Cys	HN	3.55	6.00
	HA	73 Tyr	HN	3.50	7.00
	HN	73 Tyr	HN	3.10	7.00
	HA	74 Cys	HN	4.64	6.00
	HB2	73 Tyr	HN	5.50	3.00
	HB3	73 Tyr	HN	5.50	3.00
	QB	75 Ala	HN	6.92	2.00
73 Tyr	HN	73 Tyr	HA	2.74	5.00
	HN	73 Tyr	HB2	3.76	5.00
	HN	73 Tyr	HB3	3.76	5.00
	HN	73 Tyr	QB	3.46	5.00
	HN	74 Cys	HN	3.36	6.00
	HA	74 Cys	HN	3.48	6.00
	HB2	74 Cys	HN	3.73	6.00
	HB2	75 Ala	HN	4.42	5.00
	HB3	74 Cys	HN	3.73	6.00
	HB3	75 Ala	HN	3.42	5.00
	QB	74 Cys	HN	4.28	5.00
	QB	75 Ala	HN	5.00	5.00
	QD	75 Ala	HN	7.64	3.00
74 Cys	HN	75 Ala	HN	4.07	6.00
	HA	75 Ala	HN	2.49	6.00
	QB	75 Ala	HN	4.08	6.00

77 Leu	HN	77 Leu	HA	2.40	5.00
	HN	77 Leu	QB	3.46	5.00
	HN	77 Leu	HG	4.69	5.00
	HN	77 Leu	QD1	6.53	4.00
	HN	77 Leu	QD2	6.53	4.00
	HN	77 Leu	QQD	6.39	4.00
	QQD	78 Lys	HN	6.45	4.00
78 Lys	HN	78 Lys	HB2	3.02	5.00
	HN	78 Lys	HB3	3.02	5.00
	HN	78 Lys	QB	3.00	5.00
	HN	78 Lys	HG2	5.00	5.00
	HN	78 Lys	HG3	5.00	5.00
	HN	78 Lys	QG	5.00	5.00
80 Ala	HN	80 Ala	HA	2.71	5.00
81 Lys	HN	81 Lys	QG	6.31	3.00
	HN	82 ser	HA	3.39	6.00
	QB	82 Ser	HN	6.00	3.00
	QG	82 ser	HN	6.38	3.00
82 Ser	HN	82 Ser	HB2	3.11	5.00
	HN	82 Ser	HB3	3.11	5.00

APPENDIX 5

List of Original Pulbications

Laajoki, L.G., LeBreton, E., Shooter, G.K., Wallace, J.C., Francis, G.L., Carver, J.A. and Keniry, M.A. (1997) "Secondary structure determination of ^{15}N-labelled human Long-[Arg^{3}]-Insulin-like growth factor-I by multidimensional NMR Spectroscopy" FEBS Lett. 420, 87-102

Laajoki, L.G., Francis, G.L., Wallace, J.C., Carver, J.A. and Keniry, M.A. (2000) "Solution structure and backbone dynamics of Long-[Arg^{3}]-Insulin-like growth factor-I " *J. Biol. Chem.* **275**, 10009-10015

Laajoki, L.G., Francis, G.L., Milner, S.J., Carver, J.A. and Keniry, M.A. "Defining the structural binding epitope of Insulin-like growth factor binding protein-2 on Long-[Leu^{60}]-Insulin-like growth factor-I by NMR spectroscopy" submitted to J. Mol. Biol. July 2000